

Faculty of Science
University of Helsinki

THE LUNATICS HAVE TAKEN OVER THE ASSESSMENT

UTILISING SUMMATIVE SELF-ASSESSMENT TO
THEORISE – AND DISRUPT – THE INTERPLAY OF
AGENCY AND POWER IN UNDERGRADUATE
MATHEMATICS EDUCATION

Juuso Henrik Nieminen

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Science of the University of Helsinki, in Auditorium II, Metsätalo, on the 21st of August, 2020 at 12 o'clock.

Helsinki 2020

University of Helsinki
Faculty of Science
Department of Mathematics and Statistics

Supervisors

Professor Juha Oikkonen
University of Helsinki

PhD Johanna Rämö
University of Helsinki

PhD Laura Tuohilampi
University of New South Wales

PhD Henri Pesonen
University of Helsinki

Pre-examiners

Professor Paola Valero
University of Stockholm

Associate Professor Kelvin Tan
National Institute of Education Singapore

Opponent

Professor Margaret Bearman
Deakin University

Custos

Professor Mats Gyllenberg
University of helsinki

Cover illustration

Ilona Puska

ISBN 978-951-51-6365-3 (pbk.)
ISBN 978-951-51-6366-0 (PDF)

Unigrafia
Helsinki 2020

ABSTRACT

This doctoral thesis adds to the theoretical understanding of the interplay of agency and power in self-assessment in the context of undergraduate mathematics education. This is achieved by utilising the Foucauldian notion of subject positioning, referring to the positions that assessment constructs for students. This doctoral thesis addresses summative self-assessment that involves the element of self-grading, and the disruptive nature of this practice in particular. The four substudies of this doctoral thesis investigate the reflective space that summative self-assessment opens for students to renegotiate their positioning of “the assessee” in the examination-driven context of undergraduate mathematics. Drawing on theoretical and methodological triangulation, this empirical study deepens and disrupts the interplay between agency and power in self-assessment through both positivist and socio-cultural approaches. What is utilised is the concept of resistance that highlights the importance of concrete tools for students’ agentic repositioning. Overall, the findings underline the potential of disruptive self-assessment practices in undergraduate mathematics education.

This doctoral thesis has been conducted in the Digital Self-Assessment (DISA) project, in which the summative self-assessment model was created as a self-assessment model for large undergraduate mathematics courses. Summative self-assessment is an assessment model that includes transparent learning objectives, various forms of feedback regarding those objectives and formative self-assessment practices. At the end of the summative self-assessment model students decide their own grade. In this study, the summative self-assessment model is examined through the perspective of students, particularly from the viewpoint of how students positioned themselves after taking part in summative self-assessment.

This doctoral thesis consists of four substudies. Studies I, III and IV were based on an experimental study in which the participants in a large-scale undergraduate mathematics course were randomly divided into two groups. Half of the students attended a course exam (formative self-assessment group) and half of them self-graded themselves (summative self-assessment group); both groups took part in a formative self-assessment process. After the course, 41 students were interviewed (26 from the summative and 15 from the formative self-assessment group). Furthermore, a survey study (N = 299) was conducted after the course. The data for Study II was collected through a survey in another adaptation of the summative self-assessment model (N = 113).

Studies I and II drew on quantitative methodology to examine the quality of studying (as defined through deep and surface approaches to learning, self-efficacy and course achievement) within the summative self-assessment model to shed light on the positioning processes on a broader scale. Both

studies used profiling methods to conceptualise the educational benefits of self-assessment as varying for different student subgroups. Study I drew on latent profile analysis to investigate student subgroups in terms of deep and surface approach. Four profiles were identified and compared between the formative and summative self-assessment groups. Study II, leaning on cluster analysis, examined student subgroups after another course implementation of the summative self-assessment model. Study II qualitatively looked for instructional elements that students connected with a deep approach to learning. Both studies connected summative self-assessment with a deep approach to learning, while Study I also identified a connection with a higher reported level of self-efficacy.

Study III investigated the summative self-assessment model through the concept of student agency, aiming to understand the affordances that the self-assessment model offers for agentic learning and studying. This study utilised the socio-cultural framework of ecological agency. The findings of Study III implied that the summative self-assessment model was connected with future-driven agentic behavior. Study IV introduced three different theoretical frameworks for power (sovereign, epistemological, disciplinary) to understand the socio-cultural nature of summative self-assessment as a political practice. As Study III examined pedagogical opportunities for agentic learning, Study IV sought to critically examine whether students could make use of these opportunities in spite of the complex power relations of mathematics assessment. Both studies drew on student interview data.

Finally, the findings of Studies I-IV were reinterpreted and synthesised through a discursive-deconstructive reading of these studies. What was deconstructed was students' positioning as "the assessee" and whether, and how, summative self-assessment disrupted this position. Based on the deconstruction process, the concept of transformative self-assessment was formulated. Overall, this thesis raises concerns about non-agentic subject positions that mathematics assessment tends to produce, calling for researchers to engage with disruptive practices.

Keywords: Self-assessment, agency, power, subject positioning, undergraduate mathematics education, higher education

ACKNOWLEDGEMENTS

Mathematics assessment is commonly based on the idea that knowledge and skills can be rendered as something measurable. Above all, my doctoral candidacy journey has proven this assumption about learning to be completely false. This project has truly shown that learning is not so much about the acquisition of skills as it is about becoming someone. I am not the same person I was in 2018 when I started this journey; I was never meant to become an educational researcher, yet here I am. I was *never* meant to conduct research on self-assessment - yet here we are. Another false assumption that mathematics assessment often promotes is that learning is a personal process with individual outcomes. No part of this doctoral project would have been possible without my brilliant friends, comrades and colleagues that have supported me and taught me more than I ever could have anticipated. Collaboration and communal learning can be seen throughout this doctoral thesis, and now is the time to thank everyone involved.

I owe my deepest gratitude to my supervisors who have guided me through this academic rollercoaster. First, I thank Prof. Juha Oikkonen for his support and encouragement throughout this research process. I warmly thank Johanna Rämö for her guidance and supervision at the Digital Self-Assessment Project. This whole project would not have been possible without Johanna's support and expertise. Furthermore, it is impossible to put into words how much I have learnt as a mathematics teacher having been a student of both Juha and Johanna: your role as change agents of mathematics education has inspired me a lot during my journey in academia - and beyond. I owe my deepest gratitude to Laura Tuohilampi, my dear supervisor and partner in crime (and in business!). I am thankful to Laura for having believed in me, first by inviting me to co-author a book chapter and then encouraging me to push my boundaries by starting a PhD. I am sure my adventures with Laura have only just begun. "The rockstars of mathematics education", as we were once called, will surely continue their world tour in the future! I would also like to thank Henri Pesonen for his professional guidance and sharp criticism throughout my doctoral candidacy. Henri has supported my growth as a researcher and deepened my understanding of inclusive higher education through our long conversations and collaboration. Again, I sincerely believe that our collaboration is just beginning.

I would like to express my warmest gratitude to Mira Kalalahti for her supervision on my bachelor's and master's theses before my doctoral candidacy. Mira believed in my skills and encouraged me to reach further in academia, which ultimately led me to shift my early career as a teacher. I could not write these words without Mira's professional guidance and warmth. I also greatly thank my co-authors Jokke Häsä and Henna Asikainen. Jokke has guided me in becoming the researcher I am right now; he has taught me a lot

about self-assessment, but he has also shown me how to make the most of academic conferences and their after parties. Henna has supported me in developing an academic identity in the intersection of mathematics education and higher education research. It has been a privilege to start my academic career working with such inspirational researchers.

I sincerely thank my preliminary examiners Prof. Paola Valero and Prof. Kelvin Tan for their careful reading of my work. The constructive feedback offered by both of the examiners provoked me to reflect on the research project as a whole. Conducting the final revisions based on this feedback reframed what I thought addressing feedback in academia meant, as rather than simply improving the quality of the text I ended up repositioning myself as a researcher. If my position started as the bridge-builder between various theories and methodologies, it ended up being *the disruptor of discourses*; a title I absolutely need to include in my business card from now on. Jokes aside: my heartfelt thanks for both of the pre-examiners for enabling me this final learning experience. I also express my warmest gratitude to Prof. Margaret Bearman for accepting the role of the opponent in the public defence of this thesis.

There are no words to describe how important our research group at the Department of Mathematics and Statistics has been for me during my doctoral candidacy. I speak on behalf of all of us doctoral candidates as I thank Prof. Juha Oikkonen, Johanna Rämö and Jokke Häsä for their work for establishing our mathematics education research group. We doctoral candidates have been able to actively co-create the culture of this group; it has been a privilege to grow up as a researcher through this collaborative work. I express my warmest gratitude to Juulia Lahdenperä, Jani Hannula, Jenni Honkavaara and Saara Lehto for sharing this journey with me. The four studies of my doctoral thesis were conducted under the strict surveillance and guidance of our writing group (that goes by the name *kirjoituspiiri*). Furthermore, conducting the ethical guidelines for our research group as doctoral candidates was the kind of a learning process doctoral candidates rarely have a chance to experience. Subsequently, our group has spread around Finland as the top researchers in undergraduate mathematics education - may our collaboration reign in the future as well!

The international community of PhD students in mathematics education has been a great support for me. I am sincerely thankful for the opportunity to attend the summer school by the European Society for Research in Mathematics Education in Montpellier, France, in 2018. The friendships formed at this summer school, and in other activities for young researchers, have turned out to be one of the most important outcomes of this doctoral project. I sincerely thank my dear colleagues Shu Zhang, Dorota Lembrér, and the YERME community as a whole. Let us continue our academic adventures with pride - it is us young researchers who will be the future of mathematics education research.

During this journey, I have been incredibly privileged to collaborate in a few international research projects. I am humbled to thank my dear colleagues Anette Bagger, Alexis Padilla and Prof. Paulo Tan who invited me to join their critical group of mathematics educators called New Frontiers in Mathematics Assessment. We have aimed at reaching towards those unthinkable and unnatural assessment practices that we do not even dare to speak about out loud in mathematics education. What I have learnt from all of you has greatly shifted my understanding about assessment, and the fruits of our conversations can be picked all across this thesis. Not only have I been able to openly chat with these fellow Foucauldians using the proper Foucauldian terms, but above all, I have learnt that promoting equity in mathematics education often goes hand in hand with disruption. Truly, you have shown me that these days, disruptive academic activism can appear in the form of care. I am very excited to think about our future endeavors.

What completely changed the story of my doctoral candidacy - and, as it turned out later on, my life - was a completely *lunatic* journey to Melbourne, Australia. Being a true novice when it came to academic visits, I followed my supervisor Laura for a research visit to the International Centre for Classroom Research (ICCR) at the University of Melbourne. At that point I thought that I would spend three months at the ICCR, writing my PhD articles at the office and occasionally meeting up with some new faces. I could not have been more wrong. Instead, I ended up to hiking amidst snowfall at Cradle Mountain in Tasmania and running away from deadly spiders in Sydney. What was supposed to be one single adventure to Australia led into another one, and then another; indeed, three of the four research articles that make up this thesis were mainly written in Melbourne. First, I need to thank Prof. David Clarke, Esther Chan and Carmel Mesiti from the ICCR for making these trips possible. I have rarely witnessed such hospitality in any institution. You truly made me feel like part of the community, and that feeling should not be taken for granted. I feel heartbroken that David did not have a chance to read this thesis; yet, his brilliant comments and ideas can be reflected in many pages of it. I would also like to warmly thank Prof. Gail FitzSimons for all the encouragement and conversations we have had - and for teaching me how to deal with killer snakes. The wisdom Gail has shared with me is something that will be seen in my work long after finishing this thesis. I express my warmest gratitude to Prof. Kim Beswick for arranging our research visit at the University of Tasmania. Prof. Beswick has now learnt that one needs to be careful when inviting Finns to join their birthday party on the other side of the planet. They might actually end up knocking at your door!

There is no research without funding. I warmly thank the Faculty of Science and the Department of Mathematics and Statistics at the University of Helsinki for funding two years of this journey. Also, I express my warmest gratitude to Emil Aaltonen Foundation for funding the final part of this work. I have been privileged to conduct research visits and attend conferences with

the financial help from the DONASCI doctoral school, and with the support of the Faculty of Science, University of Helsinki.

I owe my deepest gratitude to my friends who have supported me throughout my doctoral candidacy. It is a privilege to have so many inspiring teachers around me - choose your friends wisely! Thank you, Merkkü, Sanni, Anna, Hilma, Jenni, and Iida, for always being there for me. Thank you, Senni, Elsa, Tiia, and Maija, for all your support and care - and for sharing the power wherever we go together. I express my thankfulness to Saara and Mia, who diligently spent endless hours at the library with me, while we were all working on our theses. We made it!

During this PhD project, a book got written as well. Thank you, Aki, for teaching me so much about assessment during this process. I also owe my warmest gratitude to Iida and Emily, both of whom have shared my adventures in Melbourne and made them feel even more special. Let us share tacos and the dance floor again once the world has re-opened! Finally, I greatly thank Eeva for sharing the journey of doctoral candidacy with me, with all its upsides and downsides. Let us continue pondering the epistemological and ontological realms of mathematics education together!

Väitöskirjatyöskentely on johdattanut minua vuosien aikana mantereelta toiselle, mutta tämä seikkailu on oikeastaan alkanut Tikkakoskella. Kiitän suuresti perhettäni siitä, että he ovat antaneet minun kasvaa omana itsenäni ja seurata omaa reittiäni, ensin Helsinkiin ja sitten kauemmas. Kenties tärkein välietappi tällä matkalla koettiin Melbournessa. Lopuksi kiitoksen ansaitseekin kumppanini Chris. Et ole vain tukenut minua väitösaikaherruuden pyörteissä vaan antanut työlle merkityksen - sekä kiivennyt kanssani pedagogisille barrikadeille. Se, mitä olen sinulta oppinut, on kaiken arvioitavan ja vertailtavan ulottumattomissa.

CONTENTS

1	Introduction.....	13
1.1	The structure of the thesis	16
2	State of the art: Student self-assessment	19
2.1	Self-assessment, self-evaluation or self-grading?	19
2.2	The formative self-assessment model: Promoting student learning.....	20
2.3	The summative self-assessment model: Involving students in the grading process.....	21
3	How does summative self-assessment support learning and studying?.....	23
3.1	The quality of studying: Approaches to learning and self-efficacy beliefs	23
3.2	Research objective.....	24
3.3	Methodological premises	25
3.3.1	Contexts of the study.....	25
3.3.2	Research designs	25
3.4	Study I.....	30
3.4.1	Aims.....	30
3.4.2	Methods.....	30
3.4.3	Findings.....	31
3.4.4	Discussion.....	35
3.5	Study II.....	36
3.5.1	Aims.....	36
3.5.2	Methods.....	36
3.5.3	Findings.....	37
3.5.4	Discussion.....	39

3.6	A brief summary.....	40
4	Beyond the individual: Investigating agency and power	41
4.1	Student agency	41
4.2	Power.....	42
4.3	Research objective.....	45
4.4	Methodological premises	45
4.5	Study III	46
4.5.1	Aims	46
4.5.2	Methods	46
4.5.3	Findings	47
4.5.4	Discussion.....	49
4.6	Study IV.....	50
4.6.1	Aims	50
4.6.2	Methods	51
4.6.3	Findings	51
4.6.4	Discussion.....	53
4.7	A brief summary.....	54
5	Synthesis through a discursive-deconstructive reading.....	55
5.1	Summary of the methodological approaches	55
5.2	Theoretical framework: Power and subject positioning.....	57
5.2.1	Subject positions in mathematics assessment: Self-assessment as resistance	59
5.3	Discursive-deconstructive reading: Joining Foucault's game to understand self-assessment	61
5.4	The objective and process of the synthesis	62
5.5	The deconstruction of student positioning in Studies I-IV ...	63
5.5.1	Disrupting the position of the assessee.....	63

5.5.2	The disruptive practices that enabled agentic positioning.....	64
5.5.3	Maintaining the position of the assessee: What hindered agentic positioning?	65
5.5.4	Student positioning in the research project.....	67
6	Beyond deep shift: Transformative self-assessment	70
7	Self-reflection of the researcher	73
7.1	My position as a researcher, and how it has been disrupted	73
7.2	Transferability of the findings	74
7.3	Limitations and ethical considerations	75
8	Implications	78
8.1	Implication for theory	78
8.2	Implications for methodology.....	79
8.3	Implications for practice.....	80
9	Final conclusions	82

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

- I Nieminen, J. H., Asikainen, H., & Rämö, J. (2019). Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models. *Studies in Higher Education*.
doi: [10.1080/03075079.2019.1688282](https://doi.org/10.1080/03075079.2019.1688282)

- II Nieminen, J. H., Häsä, J., Rämö, J., & Tuohilampi, L. (2018). Replacing exam with self-assessment: Reflection-centred learning environment as a tool to promote deep learning. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Ed.), *Proceedings of the 21th Annual Conference on Research in Undergraduate Mathematics Education*. San Diego, CA: Special Interest Group of the Mathematics Association of America for Research in Undergraduate Mathematics Education.

- III Nieminen, J. H. & Tuohilampi, L. (2020). ‘Finally studying for myself!’ Examining student agency in summative and formative models of self-assessment. *Assessment & Evaluation in Higher Education*. doi: [10.1080/02602938.2020.1720595](https://doi.org/10.1080/02602938.2020.1720595)

- IV Nieminen, J. H. (2020). Disrupting the power relations of grading in higher education through summative self-assessment. *Teaching in Higher Education*.
doi: [10.1080/13562517.2020.1753687](https://doi.org/10.1080/13562517.2020.1753687)

The publications are referred to in the text by their roman numerals.

1 INTRODUCTION

“Gracious heavens - the lunatics have most undoubtedly broken loose!”

In Edgar Allan Poe’s short story “The System of Doctor Tarr and Professor Fether” (1845), an unnamed doctor visits a famous mental institution in southern France. The institution is known for its groundbreaking methodology for treating patients. All punishments are avoided, and the patients are not confined to cells but treated as a community - in fact, the word “lunacy” is not employed at all. The visiting doctor soon finds out from the superintendent, however, that this revolutionary treatment method has been abandoned. Curious to hear the rationale behind this decision, he joins a dinner party with the superintendent’s unusual colleagues. They behave strangely and make unsettling jokes about the patients, such as one who “very pertinaciously maintained himself to be a Cordova cheese”. They tell the doctor that the institution now uses a different system, invented by the famous “Doctor Tarr and Professor Fether”. The doctor has never heard of them - even though he is an educated man! The superintendent explains that this new model is a very strict one; it was implemented in response to a violent uprising during which the patients attempted to create a lunatic government. As the dinner progresses, the shocked doctor learns that his fellow dinner guests are in fact “the lunatics”. The real doctors have been locked up as madmen, and been tarred and feathered in the cells to keep them from causing harm. It turns out that their positions are now completely changed.

This doctoral study examines self-assessment and its interplay with agency and power in the context of undergraduate mathematics. Just as Poe’s story questioned the positions in a mental institute, this thesis disrupts the positions that mathematics assessment, and self-assessment in particular, construct for students. Teacher-driven assessment practices have been claimed to mainly position students as assesseees, the receivers of assessment (e.g., Torrance, 2000). Alternative practices, such as the one of summative self-assessment, might offer students opportunities to position themselves differently. This doctoral thesis critically examines that act of empowerment through self-assessment by examining summative self-assessment that allows students to award their own grades. The Poe-inspired title is a refrain all-too-familiar to the developers of the assessment model in question: “the lunatics have taken over the assessment - this is not how you assess mathematics!” Like the groundbreaking treatment system in Poe’s story, the summative self-assessment model aims to foster agency. In the story, the novel methodology had ironic consequences; this doctoral thesis investigates whether promoting student agency through summative self-assessment would lead to a similar aftermath.

The premises of this doctoral thesis are rooted in the socio-cultural end of the higher educational Assessment for Learning movement (William, 2011). The goal for higher educational institutes has been to educate students to critically engage in and take responsibility for their own learning rather than to memorise fragmented pieces of information. The contribution of assessment has been increasingly noted in this process (see Boud, 2007; Boud and Falchikov, 2006; Falchikov, 2005). At the same time, summative assessment methods are largely teacher-led. It has been argued that grading processes dominate learning in neoliberal higher education (e.g., Torrance, 2000, 2007) and that students would be seen merely as the subjects of assessment; “they are recipients of the actions of others, not active agents in the assessment process” (Boud & Falchikov, 2006, p. 403; see also Evans, 2011). These concerns seem crucial in undergraduate mathematics education where assessment is mostly conducted through examinations (Iannone & Simpson, 2011).

This doctoral thesis develops earlier work on self-assessment and power in the field of higher education (e.g., Leach, Neutze, & Zepke, 2001; Tan, 2007, 2008; Taras, 2001, 2008, 2016). However, rather than focusing on the theory itself, this doctoral thesis offers empirical evidence on whether this alternative self-assessment practice would promote student agency within the power relations of undergraduate mathematics assessment. Earlier research on summative self-assessment in higher education is scarce - a recent literature review (Andrade, 2019) only identified one study that drew on self-grading. Also, as previous studies on self-assessment have underconceptualised the notion of student agency (Bourke, 2018; Milne, 2009), this doctoral thesis aims to theorise the interconnection of self-assessment and agency by tying it with the notion of power.

This doctoral thesis consists of four studies that each contribute to understanding the interconnection of agency, power and summative self-assessment. The datasets were collected from two large undergraduate mathematics courses that utilised the summative self-assessment model. Hence, students practiced self-assessment during the courses, and at the end of it they were responsible for choosing their own grade. After the courses, data on the quality of students’ studying was collected through a survey. Furthermore, interviews were conducted to map out student perceptions of summative self-assessment. The four studies are based on both theoretical and methodological triangulation (Denzin, 1978; Van Drie & Dekker, 2013) as the studies do not just use both quantitative and qualitative methods but also share different epistemological premises. As will be shown, the ruptures between these different approaches and premises cannot always be bridged, but the ruptures themselves act as important research findings. The first two studies, Studies I and II, draw on quantitative profiling analyses to identify student subgroups in terms of deep and surface approaches to learning (e.g., Entwistle & Ramsden, 1983; Marton & Säljö, 1976). In Study II, students’ open answers in the survey are also analysed to categorise their perceptions of self-

assessment. Studies III and IV concern the interview data as collected after one of the two courses. Study III examines the affordances that the summative self-assessment model offers for student agency (Emirbayer & Mische, 1998; Biesta & Tedder, 2007), while Study IV focuses on those complex power relations that control the use of those affordances (e.g., Tan, 2007, 2008).

The four studies of this doctoral thesis are synthesised through a discursive-deconstructive reading (Ikävalko & Brunila, 2019) that reconceptualises the findings of Studies I-IV through a discursive lens by drawing on Foucault's notion of subject positioning (Foucault, 1977, 1982, 1991). This doctoral thesis examines the mechanisms of subject positioning connected with summative self-assessment in undergraduate mathematics, and how agency and power play their parts in the process. As previous studies have mainly investigated the position of the assessee for students in higher education (e.g., Boud & Falchikov, 2006; Evans, 2011), summative self-assessment aims to disrupt this positioning and offer students opportunities for agentic repositioning. Whether the self-assessment model disrupts the position of the assessee is discussed in this thesis. In Poe's story, various characters were positioned by other actors, yet the most crucial aspect was the self-positioning of the "lunatics" who questioned the mere concept of madness. This doctoral thesis addresses similar kind of agentic self-positioning in terms of summative self-assessment.

Poe's story focuses on another theme important to this doctoral thesis: resistance. Enabling students to have power over their grades, namely through self-grading, is strictly not recommended by the previous literature. For example, Andrade and Valtcheva (2009, p. 17) warn about self-grading: "Do not turn self-assessment into self-evaluation by counting it toward a grade" (see also Andrade, 2019; Bourke, 2018). At the same time, self-assessment is described as a practice that would promote student agency (Bourke, 2018; Milne, 2009; Taras, 2016). Earlier research, however, has not elaborated on what exactly the interaction between self-assessment and agency is. Here, summative self-assessment is introduced as a method to study the usual subject positions in undergraduate mathematics assessment by causing "a breach of self-evidence" (Foucault, 1991, p. 76). At the same time, there is a political, rather than apolitical, goal of resistance. As a practice, summative self-assessment aims to foster critical reflection and asks students to learn for themselves, not for their teacher.

Supplementing the earlier literature on subject positioning in assessment that have drawn on theoretical reviews (Taras, 2001, 2008, 2016) and higher educational documents (Evans, 2011), this doctoral thesis focuses on the perspective of the students. This perspective seems fitting in relation to self-assessment as students themselves "dominate the whole process and their internal values, ideas, goals and skills are extremely important" (Yan & Brown, 2017, p. 1248). Also, as Tan (2004) notes, it is not only teachers but students as well who bring their adopted positions into the assessment process. This makes undergraduate mathematics education a particularly interesting setting

for the study, as earlier studies have confirmed that students prefer their assessment to be based on traditional methods such as examinations in this context (Iannone & Simpson, 2015a, 2015b). However, following Raaper (2019), this doctoral thesis does not see students as non-agentic recipients of the power relations but active negotiators whose experiences and insights affect these power relations.

Even though this doctoral study introduces many theoretical frameworks (and ultimately groups them under a final one), in its core it is based on the practical issue of how to develop assessment in undergraduate mathematics education - and for what reasons. That was the starting point for the Digital Self-Assessment (DISA) project in which the study was conducted. It is argued that assessment practices in the contexts of mathematics and higher education are at risk of hindering the development of student agency if theoretical understanding of that concept is not deepened (see Charteris & Smardon, 2018). I position this doctoral thesis as a pedagogical one, as the theoretical framework of subject positioning is utilised by joining Foucault's game (Foucault, 1991, p. 74) to apply his ideas into practice rather than focusing on those theories per se. It will be argued that this element of practically disrupting the existing power relations is a substantial part of resistance (cf. Allen, 2011).

In Poe's story, the dilemma was to enable the patients the right amount of agency - too much agency led to a revolution. Is this an existing threat in undergraduate mathematics assessment? Maybe radical self-assessment methods would not get teachers tarred and feathered, but "the notorious 'I give myself an A'" (Andrade & Du, 2007, p. 160) might act as the pedagogical equivalent. It will be argued that enabling the possibility of agentic studying through summative self-assessment opens the door not only for a higher quality of studying, but for future-oriented agency as well. This reflects the stated purpose of Finnish higher education to educate students to "serve their country and humanity at large" (Finnish Universities Act, 558/2009) rather than teaching them an accustomed set of skills. Could such an idea really be so radical?

1.1 THE STRUCTURE OF THE THESIS

In this doctoral thesis, summative self-assessment was studied through through three phases. Each of which consisted of their own research objective (Table 1). These phases differed not only in terms of their methodologies, but they also drew on different epistemological premises; this is why they are presented separately. Each phase consists of a rather traditional structure for reporting scientific research with their theoretical backgrounds, methodology sections, research findings, and so on.

After introducing the state of the art concerning self-assessment literature, Studies I-II and Studies III-IV are presented separately in two phases. This

choice represents the epistemological rupture between these two phases, as Studies I and II draw on a positivist approach and Studies III and IV on socio-cultural premises. Finally, in the third phase, the four studies are synthesised through a discursive approach. Furthermore, based on the synthesis, the concept of transformative self-assessment is constituted in this final phase. This structure represents the role of each of these phases. Studies I-II builds a broader picture of students' learning and studying while taking part in summative self-assessment, while Studies III-IV deepen these findings through a qualitative investigation of agency and power based on student interviews. Finally, these two phases are synthesised through the theoretical framework of subject positioning that deepens - and disrupts - the theoretical understanding of agency and power in terms of the research process.

The structure of this doctoral thesis reflects the chronological order of the PhD project as a whole. This choice highlights the growth of the researcher during the process, as my own position as a certain kind of researcher was disrupted during the project. This is why the research objective for the whole project consists of three separate research objectives (Table 1) that evolved throughout the process and whose connections will be discussed. The structure represents not only the smooth connections between the four studies but the ruptures included in the research process. This choice fits the main argument of the thesis: that as a practice, self-assessment itself disrupts rather than reconciles. It should be noted that the model of summative self-assessment is introduced in Section 3.3.2.

Table 1. *The structure of the doctoral thesis.*

	Epistemological premises	Main concepts	Research objective
Studies I & II	Positivist	Approaches to learning, self-efficacy	Based on quantitative and mixed methods approaches, how is summative self-assessment connected with deep and surface approaches to learning and self-efficacy?
Studies III & IV	Socio-cultural	Agency, power	Based on a qualitative approach, what kind of affordances does summative self-assessment offer for agentic studying and what kind of power relations affect the use of these affordances?
Synthesis	Discursive	Subject positioning	Deepening the theoretical understanding of agency and power in terms of summative self-assessment: What kind of subject positioning processes are identified in terms of Studies I-IV?

2 STATE OF THE ART: STUDENT SELF-ASSESSMENT

This section interrogates the notion of self-assessment before elaborating on the thesis' theoretical frameworks. First, various conceptualisations of student self-assessment are introduced. Next, the two self-assessment models studied and developed in this doctoral thesis and their connection with self-assessment literature are showcased: the formative and the summative models of self-assessment. A brief synthesis of self-assessment literature in the contexts of mathematics education and higher education is introduced to understand the role of self-assessment in undergraduate mathematics education.

2.1 SELF-ASSESSMENT, SELF-EVALUATION OR SELF-GRADING?

What exactly is meant by the concept of student self-assessment? Even though self-assessment is often introduced as an individual assessment method, the concept refers to a vast collection of practices (Andrade, 2019; Andrade & Du, 2007). In their review article in the context of K-16 education, Brown and Harris define self-assessment as a “descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities” (p. 368). Panadero and colleagues (2016) see it as a “wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products” (p. 804). These decontextualised definitions seem broad; indeed, Andrade suggests that they cover “everything but the kitchen sink” (2019, p. 2). Both these recent definitions include an important element of student self-assessment in educational settings: students need to engage in a process of internalising learning objectives and then make evaluative judgments based on them (Boud & Falchikov, 1989; Yan & Brown, 2017). These kinds of skills are often advocated as valuable for life-long learning in general (Boud, 2007).

For the purposes of this doctoral thesis, Andrade and Du's (2007) typology of different educational self-assessment practices is utilised. They distinguish between self-assessment, self-reflection and self-evaluation. It is important to highlight the difference of these concepts in the context of Finland, where the word *itsearviointi* could refer to any one of them. In English, *self-reflection* refers to the internal, psychological process during which the student investigates their own general qualities, attitudes and dispositions (Yan & Brown, 2017). In educational institutes, students are often asked to self-reflect through *self-assessment*, that is usually aimed towards a certain product (e.g.,

a mathematical task) or the process (e.g., mathematical strategies). Self-assessment requires a set of criteria (Andrade & Brown, 2016; Panadero, Tapia, & Huertas, 2012); this distinguishes it from those daily life situations during which students observe their own behavior. Finally, by the concept of self-evaluation Andrade and Du (2007) refer to students' judgments about their own learning after the learning process.

As Andrade (2019) argues, "any definition of self-assessment must acknowledge and distinguish between formative and summative forms of it" (p. 2). In this doctoral thesis the element of whether self-assessment is used as a formative element during the learning process (e.g., Black & Wiliam, 2003) or as a summative method after the learning process (Knight, 2002) is indeed considered. Next, the two models studied in this doctoral thesis are introduced through the concepts of formative and summative self-assessment. It is notable that, as will be argued, these two models differ in their pedagogical purpose rather than just through practicalities.

2.2 THE FORMATIVE SELF-ASSESSMENT MODEL: PROMOTING STUDENT LEARNING

Formative self-assessment builds on the idea that students would engage in self-assessment tasks during their learning process by reflecting on their own learning based on a pre-set criterion for learning (Andrade, 2010; Andrade & Du, 2007; Brown & Harris, 2013; Panadero et al., 2016). Therefore, the idea of formative self-assessment reflects the concept of self-assessment by Andrade and Du (2007). It is notable that formative assessment often includes the idea that summative assessment - namely, grading - would be teacher-driven. The pedagogical purpose of formative self-assessment is to support learning and make it visible, which is why it is often recommended to be used in addition to more traditional assessment practices rather than replacing them (Andrade, 2019). Formative self-assessment excludes students from the grading processes, but as Yan and Brown (2017) note, it places the emphasis on the students' point of view:

Thus, in self-assessment, students dominate the whole process and their internal values, ideas, goals and skills are extremely important – -. While formal or structured self-assessments are initiated and designed in educational settings by the teacher or the curriculum, the process of self-assessment is still conducted and monitored by students themselves. (p. 1248)

Brown and Harris (2014) frame self-assessment as a key competence rather than as an assessment method. This kind of a view is commonly shared in self-assessment research, as practicing self-assessment skills is often raised up as an important feature of formative self-assessment (Panadero et al., 2016). Andrade (2019) emphasises the same, underlining the importance for

students to practice formative self-assessment and to make corrections based on it: “If there is no opportunity for adjustment and correction, self-assessment is almost pointless” (p. 2). Both internal and external feedback are important factors in the self-assessment process (Yan & Brown, 2017); feedback should be offered both on the content and on students’ self-assessment skills. Building on these views, formative self-assessment fits with the idea of feedback cycles and spirals (Beaumont, O’Doherty, & Shannon, 2011; Carless, 2019) during which self-assessment is only one of the sources for continuous feedback that the students can act on. Through formative self-assessment, students learn to calibrate their ideas about their own skills and knowledge in relation to the learning objectives (Panadero et al., 2016; Yan & Brown, 2017). This kind of formative self-assessment process has been repeatedly connected with increased learning results and better quality of learning (for literature reviews see Andrade, 2019; Panadero et al., 2016).

Why is formative self-assessment promoted without letting students take part in the grading process? Often, concerns about the validity of self-assessment are raised (e.g., Brown, Andrade, & Chen, 2015). However, what is lacking in the higher educational research of self-assessment literature is empirical evidence on student behavior in relation to self-grading - and an elaboration on the socio-cultural aspects related to self-grading. Despite this, the message of earlier literature is clear: “Do not turn self-assessment into self-evaluation by counting it toward a grade” (Andrade & Valtcheva 2009, 17). Bourke (2018) argues that in higher education, self-grading would lead into a focus on grades rather than on learning, yet offers no empirical data or scientific references to support this statement. It has even been suggested that “human nature” might cause dishonesty during self-grading (Andrade, 2010, p. 92). To conclude, warnings about keeping self-assessment formative have been common, and formative self-assessment is by far the most common model for educational purposes; for example, in a recent literature review (Andrade, 2019), only one higher educational study was identified involving self-grading (the study by Tejeiro et al., 2012).

2.3 THE SUMMATIVE SELF-ASSESSMENT MODEL: INVOLVING STUDENTS IN THE GRADING PROCESS

The summative model for self-assessment is not exclusively different in relation to the formative one; rather, it builds on it. Summative self-assessment includes all the same elements as the formative model. Yet, summative self-assessment does not just allow students to compare their work against certain criteria but also involves students in the summative assessment process by letting them assign their own grade (Strong, Davis, & Hawks, 2004; Taras, 2001, 2016, 2008). Therefore, summative self-assessment includes not only the element of self-assessment but the one of self-grading (Andrade & Du, 2007) as well. As the model builds on the formative one with its iterated

feedback processes (Beaumont, O'Doherty, & Shannon, 2011; Carless, 2019), it fosters active engagement with the self-assessment process rather than simply letting students grade themselves at the end of the learning process. Summative self-assessment can be seen as a "process within a process, in which many thoughtful and fair decisions have to be made according to pre-established and reasonably set criteria" (López-Pastor et al., 2012, p. 454).

Thus, the summative self-assessment model does not add the element of self-grading on top of the formative model. On the contrary, summative self-assessment aims to reconceptualise the pedagogical purpose of self-assessment by tying assessment with the notion of *future-driven assessment* as introduced by Tan (2008, 2007). Future-driven self-assessment aims to develop skills of lifelong learning that would be needed outside university courses or higher educational programmes. As future-driven self-assessment emphasises "students' capacity for exercising their own judgements without depending on the academic" (Tan, 2007, p. 119-120), summative self-assessment connects with the idea by showing students that they are now responsible for their own learning - yet only through a scaffolded process. Summative self-assessment aims to tackle the issue that Tan (2007) reports about students trying to please their teacher or trying to adjust to the needs of their programmes.

A similar view is shared by Boud and Falchikov (2006) who argue that students need to be seen as active agents in their assessment and learning processes since "neither teacher or the curriculum drives learning after graduation" (p. 402). This is not to say that formative self-assessment could not teach these skills. However, the pedagogical purpose of summative self-assessment is future-driven rather than teacher-driven (Tan, 2007); for example, feedback provided by the teacher during summative self-assessment is only offered as a base for further reflection, while the students themselves have the power to evaluate whether they have reached the learning objectives for the grade they claim (Taras 2001, 2016, 2008).

3 HOW DOES SUMMATIVE SELF-ASSESSMENT SUPPORT LEARNING AND STUDYING?

In this section, Studies I and II are introduced and discussed. These studies started the process of conducting this doctoral thesis. Based on earlier literature, it could be argued that summative self-assessment might support students' learning and studying. The aim of the Digital Self-Assessment (DISA) project was to empirically investigate whether this held true in practice; whether the summative self-assessment model supported learning. Therefore, two survey studies were designed to measure students' quality of studying and course achievement at two separate undergraduate mathematics courses. The approach taken was deliberately psychological. Next, the theoretical framework of Studies I and II is presented, followed by the introduction of these two original studies.

3.1 THE QUALITY OF STUDYING: APPROACHES TO LEARNING AND SELF-EFFICACY BELIEFS

Studies I and II utilised the theoretical frameworks of approaches to learning (e.g., Asikainen & Gijbels, 2017; Entwistle & Ramsden, 1983; Marton & Säljö, 1976) and self-efficacy (e.g., Bandura, 1997, 2000) to quantitatively operationalise the quality of studying. The *student approaches to learning* tradition has been widely used in the field of higher education to examine how students study in various contexts (Entwistle & Ramsden, 1983; Marton & Säljö, 1976; Biggs, 1991). Traditionally, these approaches have been divided into deep and surface approaches to learning; where a deep approach emphasises students' aim to understand the content through critical thinking, whereas a surface approach refers to memorising the content through rote learning (Entwistle & Ramsden, 1983). Research has repeatedly connected a deep approach to learning with higher achievement in higher education (Biggs, 1991; Diseth, 2003; for the context of undergraduate mathematics, see Lahdenperä, Postareff, & Rämö, 2019; Maciejewski & Merchant, 2016; Murphy, 2017).

Two features of this tradition should be highlighted in relation to this doctoral thesis. First, the student approaches to learning tradition is situational, meaning that students' approaches only exist in relation to their specific learning environments (e.g., Entwistle & Ramsden, 1983). This separates the tradition from motivation theories, for example, since students might utilise a deep approach in one context and surface approach in another. Also, students often use a combination of these two approaches rather than only drawing on one of them (e.g., Parpala et al., 2010).

As approaches to learning are situational and therefore able to change, a voluminous body of research has observed the ways that teaching and assessment could support a deep approach and prevent a surface approach to learning. It has been claimed that assessment has a huge impact on approaches to learning (Rust, O'Donovan, & Price, 2005). But how exactly could assessment support students' "deep shift" (Wilson & Fowler, 2005)? Earlier research has shown that alternative forms of assessment such as self- and peer-assessment might actually even promote a surface approach (Gijbels & Dochy, 2006) or, at best, prevent students from applying a surface approach to learning (Baeten, Dochy, & Struyven, 2008; Struyven et al., 2006). Haggis (2003) raised the question as to whether a deep approach to learning could even be "induced" if it is not "already there" (p. 94) – this seems to ring true in the field of assessment in particular. As research on the connection between self-assessment and approaches to learning is scarce, this doctoral thesis offers new empirical evidence on how two different self-assessment models could support a deep approach and prevent a surface approach to learning (conceptualised through the notions of agency and power).

To supplement the framework of approaches to learning, this doctoral thesis also utilises the widely used concept of self-efficacy to operationalise student agency. Generally, students' self-efficacy beliefs can be defined as a person's belief about their own abilities to achieve in a given form of attainment (Bandura 1989, 1997, 2000). Therefore, as is the case for approaches to learning, self-efficacy beliefs are situational (as compared to concepts such as *self-esteem* or *self-image*). High self-efficacy beliefs have been widely connected with greater achievement (for a meta-analysis, see Richardson, Abraham, & Bond, 2012) and with a deeper approach and lower surface approach to learning (Prat-Sala & Redford, 2010). Furthermore, the positive relationship of self-assessment to higher self-efficacy beliefs has been shown in previous studies (e.g., Panadero, Jönsson, & Botella, 2017). In their four (!) meta-analyses, Panadero and colleagues (2017) suggested that this might be due to self-assessment teaching the student valuable information about the requirements of a specific task, which leads to successful performance. This idea is in line with Bandura's (1997) finding that self-efficacy beliefs can be developed through experiences of mastery; as students gain feelings of mastery in self-assessment, their self-efficacy beliefs might also be promoted.

3.2 RESEARCH OBJECTIVE

The overall objective for Studies I and II was to investigate the quality of learning (deep and surface approaches to learning) and self-efficacy in terms of both formative and summative self-assessment. These two studies were based on two separate undergraduate mathematics courses, of which Study I drew on both formative and summative self-assessment and Study II on

summative self-assessment. Furthermore, both studies drew on profiling as they focused on studying student subgroups rather than the student population as a whole; the underlying assumption was that not all students would benefit from self-assessment in similar ways. Also, as earlier studies (e.g., Ibabe & Jauregizar, 2010; Jay & Owen, 2016) have suggested that self-assessment relates to higher achievement through students' active engagement in their own learning process, academic performance was investigated in both Studies I and II.

3.3 METHODOLOGICAL PREMISES

3.3.1 CONTEXTS OF THE STUDY

As this doctoral thesis as a whole is based on a socio-cultural approach which conceptualises assessment as situational and associated with social practices, several contexts should be emphasised. The broader context to the study is Finnish higher education, where grades in general do not determine students' educational paths. Examinations can often be taken multiple times. In keeping with the Finnish Universities Act (2009), teachers have autonomy over their teaching and assessment methods; Finland scores very high internationally on academic freedom (Nokkala & Bladh, 2014). It should be emphasised that the Finnish context offers a fertile ground for assessment experiments such as the one reported here.

Undergraduate mathematics, another context of this doctoral study, has been shown to be an examination-driven culture in which students want to be assessed through traditional methods (Iannone & Simpson, 2015a). In Finland, no studies have investigated *how* assessment is usually conducted in undergraduate mathematics. However, a recent Finnish report highlighted that at the secondary and basic levels of education, mathematics is commonly assessed through traditional assessment methods such as individual examinations (Atjonen et al., 2019). In the same national report it was found that, according to the teachers, STEM subjects, with mathematics being a part of them, scored the lowest of all school subjects in the use of self- and peer-assessment. Even though the present study is positioned in the context of higher education, the report by Atjonen and colleagues characterises the culture of mathematics assessment in Finland and depicts the general assessment environment relevant to the examinees of this study.

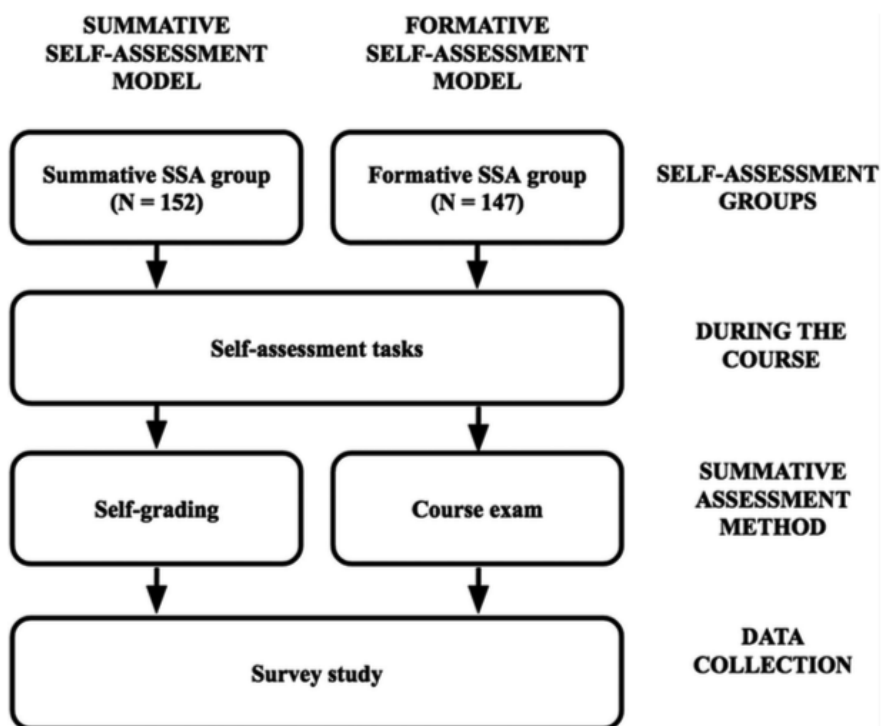
3.3.2 RESEARCH DESIGNS

Studies I and II were based on different implementations of the summative self-assessment model; the research designs themselves differed as well. Here,

the self-assessment model implementation and research design of Study I is introduced first, and that of Study II follows in terms of how it differed from Study I. Studies III and IV were based on the same course adaptation as Study I. The concepts defined in this section (such as self-assessment *models* and self-assessment *groups*) are used throughout this doctoral thesis. All of the studies were conducted as a part of the Digital Self-Assessment (DISA) project at the University of Helsinki.

For Study I, an undergraduate mathematics course in a research-intensive university in Finland was investigated (see Figure 1). The 5 credit (European Credit Transfer and Accumulation System) course lasted for seven weeks. There were 426 participants at the beginning of the course, of which 313 were actively engaged and passed it; of these, 299 participated in Study I. The topic of the course was linear algebra; it is one of the first courses mathematics students take, covering topics such as systems of linear equations, vectors and matrix algebra. Overall, the course was designed to be student-centred. Teaching was based on the Extreme Apprenticeship Model (for details, see Rämö, Reinholz, Häsä, & Lahdenperä, 2019). It is a teaching model based on Flipped Learning. The Moodle online learning environment was used during the course. The course was graded on a scale from 0 (“fail”) to 5.

Figure 1 An overview of the design of Study I. The summative and formative models only differed in terms of their final, summative grading method.



At the beginning of the course, the participants were randomly divided into two groups and informed about their placement. Half of the students attended a course examination at the end of the course (*formative self-assessment group*, studying with the formative self-assessment model), while the other half self-graded themselves (*summative self-assessment group*, studying with the summative self-assessment model). Both groups practiced self-assessment during the course as both models emphasise active engagement with the self-assessment process. Also, both groups were motivated to self-assess as a result of lecturers telling them that learning how to evaluate one's own work is an important skill and that they, the students, should use the opportunity to learn for themselves, and not just for the teacher. It is notable that only the final summative assessment method was different for the two groups; otherwise, both groups experienced exactly the same learning environment. Finally, after the final summative assessment, the data collection was conducted with a survey. In the following section, it is explained how the two self-assessment models were implemented in practice in Study I (Figure 1). Finally, the summative self-assessment model implementation of Study II is introduced in terms of how it differed from the model of Study I.

3.3.2.1 The formative self-assessment model in practice

The students in the formative self-assessment group completed self-assessment tasks during the course; however, these self-assessments did not count towards their grade. The final summative assessment was a course examination. To support students' self-assessment, the course utilised a detailed rubric to communicate the learning objectives. Some topics in the rubric were content-specific, such as "solving linear systems", while others concerned generic skills, such as "reading and writing mathematics". Examples of the learning objectives are given in Table 21. Of the topics, five concerned mathematical content and four concerned generic skills. The criteria were given at three levels, for grades 1–2, 3–4, and 5.

The students completed two compulsory self-assessment tasks during the course. In the first task, the students were shown all the learning objectives that they had worked on so far. For each objective, they stated whether they felt they mastered it: (1) well, (2) partially, or (3) not yet. Also, by using scripts (Panadero, Tapia, & Huertas, 2012), the students were asked to reflect in writing on how they thought they had mastered the learning objectives and what goals they had for the rest of the course. In the second self-assessment task, the students had to decide what grade they would award themselves for each topic in the rubric. Again, questions concerning the students' feelings and goals were asked. Also, the students had a chance to justify their self-assessment for each of the learning objectives in writing.

¹ The rubric can be accessed online in bit.ly/LinearAlgebraRubric.

The course largely utilised feedback cycles (Beaumont, O'Doherty, & Shannon, 2011) to support students' formative self-assessment. Digital feedback on students' self-assessments was offered. Each of the tasks in the course was linked with the learning objectives it was supporting and, based on the number of the tasks completed, the students received a computed index that indicated how accurately their self-assessment was in line with the work they had done during the course. It was explained to them that the indices were not necessarily representative of their skills, and they were encouraged to explain in writing if they believed that the coursework assessment did not adequately reflect their skills.

Feedback cycles were also utilised with the mathematical tasks during the course. New topics were introduced through scaffolded tasks. Each week, students were given three sets of mathematical tasks, each presenting a different kind of feedback. First, there were digital tasks offering automatic constructive feedback. These were followed by pen-and-paper tasks which were divided into two sections. The first section comprised two or three tasks concerning the most central topics of the course. One of these tasks was selected for feedback that was provided by the student tutors who were taught to write constructive feedback. Students had an opportunity to return a revised solution twice. The second section of pen-and-paper assignments consisted of tasks for which no feedback was provided; model answers for these tasks were published later.

During the course, students were offered guidance in an open drop-in learning space by student tutors who were trained in effective teaching methods (Rämö et al., 2019). The learning space enabled an opportunity for social interaction and for peer feedback. Also, digital peer assessment on mathematical tasks was provided in Moodle, and digital feedback on students' peer assessments was offered based on how constructive they were.

Table 2. *Part of the rubric of the course.*

Topic	Skills corresponding to grades		
	1-2	3-4	5
Matrices	I can perform basic matrix operations and know what zero and identity matrices are	I can check, using the definition of an inverse, whether two given matrices are each other's inverses	I can apply matrix multiplication and properties of matrices in modelling practical problems
Reading and writing	I use course's notation in my answers	In my solutions, I write complete, intelligible sentences that are readable to others	I can write proofs for claims that concern abstract or general objects

3.3.2.2 *The summative self-assessment model in practice*

The students in the summative self-assessment group took part in exactly the same learning environment as the students in the formative self-assessment group. The only difference was the final summative assessment method. Therefore, the previous description of the feedback cycles concerns this group as well. While the formative self-assessment group took the course examination, the students in the summative self-assessment group took part in a self-grading process. At the end of the course, students in the summative self-assessment group self-graded themselves in the same manner as in the second self-assessment task: grading was based on the topics in the rubric. For each grade, students could reflect on why they chose that grade, in writing. They also awarded themselves the final grade. No instructions were provided on how the summative self-assessment group should arrive at the final grade.

The digital feedback system, normally used to offer feedback on students' self-assessment, was used at the end of the course to check the self-graded marks before their final validation. This was done to ensure that students with low self-efficacy would not assess themselves with a very low grade, and also to prevent obvious cheating, as compared with the previous experiences of the teacher of the course. At the beginning of the course, all of the students were told that the validation system was only used to prevent obvious cheating and

not for reducing their power over their own grades. The system identified those students whose self-assessed and computed grades differed by more than one grade (for details, see Study I). In total, there were thirty-two such students, and their grades were dealt with separately by the teacher of the course.

Study II is the only one of the four studies of this doctoral thesis that is not based on the same course adaptation as Studies I, III and IV. Instead, it drew on a course adaptation that utilised a similar implementation of summative self-assessment model as the one described above; the course drew entirely on this summative model. The course was the same linear algebra course as in Study I, but this time the course was taught at the University of Helsinki Open University by a different teacher. The six-week course was smaller: it had 164 participants, of which 113 participated in the study. The course utilised the same rubric, and the mathematical tasks were almost identical. In comparison to Study I, the data for Study II was collected after the final summative self-assessment task.

3.4 STUDY I

3.4.1 AIMS

Study I drew on a quantitative approach to compare students' learning and studying in the summative and formative self-assessment groups. The study utilised latent profile analysis to explore student subgroups based on a questionnaire on deep and surface approach to learning ($N = 299$). Students' approaches to learning, self-efficacy and course achievement were compared in terms of the student profiles and the self-assessment groups.

3.4.2 METHODS

Study I was based on a survey study ($N = 299$; response rate 96.5 %, three students were excluded from the data as they had not answered all questions), conducted after the experimental study on formative and summative self-assessment. 152 of the participants took part in the summative self-assessment group and 147 in the formative self-assessment group. As the self-assessment groups were randomly assigned, there were no statistically significant differences between them in terms of age ($M = 24.37$, $SD = 7.02$, median = 21), major ($\chi^2(9, N = 299) = 5.18$, $p = .82$; 24 majors were represented, and 94 students majored in mathematics) or gender ($\chi^2(3, N = 299) = .35$, $p = .95$). The students signed their consent forms that their data could be used as a part of the study as they registered for the course. Also, the students taking part in

the experimental study were informed that they could withdraw from the study at any point.

Students' deep and surface approaches to learning were measured through the HowULearn questionnaire (Parpala et al., 2010; Parpala & Lindblom-Ylänne, 2012) that has been validated in the context of Finnish higher education (Herrmann, Bager-Elsbor, & Parpala, 2017), consisting of four items for both deep and surface approaches. Self-efficacy was measured with the five-item scale from the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991). Also, information about students' mathematical achievement was collected from the Moodle environment. This data concerned students' scores on the mathematical tasks during the course. The score for achievement during the course was based on the scores of the three mathematical task sets: tasks with automatic feedback, tasks with feedback from the student tutors, and tasks with no feedback. A weighted average was measured for each of the students based on these scores. As achievement was only measured through these teacher-developed tasks, the achievement score can only be seen as indicative of student learning.

The analysis of Study I was based on latent profile analysis. First, confirmatory factor analysis was conducted on the scales measuring deep and surface approaches to learning to ensure the construct validity of the instrument; the fit was found to be suitable. After deleting one item from the surface approach scale that did not fit the culture of university mathematics ("often I had to repeat things to learn them"), the fit was acceptable (CFI = .98, RMSEA = .04; Hu & Bentler, 1999). Latent profile analysis was then conducted for the whole student population to map out the latent subgroups of students in terms of deep and surface approaches to learning. Six fit indices were used to compare different profile solutions: Akaike Information Criterion (AIC; Akaike, 1987), Bayesian Information Criterion (BIC; Schwarz, 1978), the BIC Sample-Size Adjusted (aBIC), the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test and the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR LRT; Lo, Mendell, & Rubin, 2001). Also, the size of the smallest profile and the interpretability of the profile solution were considered in the analysis. Finally, the distribution of the students' profiles was compared with a Chi-square test between the two self-assessment groups.

3.4.3 FINDINGS

First, the general-level comparisons showed significant differences between the two self-assessment groups. A *t*-test analysis showed that the surface approach to learning was reported as being significantly greater in the formative self-assessment group ($t(297) = -2.5, p = .013, d = .37$), while the deep approach to learning was reported as being significantly greater in the summative self-assessment group ($t(297) = 3.26, p < .001, d = .29$). However, the effect sizes were only small or moderate. In addition, self-efficacy was

reported to be significantly higher in the summative self-assessment group with a larger effect size ($t(297) = 5.03, p < .001, d = .59$).

The latent profile analysis resulted in a four-profile solution (see Table 7). While the AIC and aBIC indexes favoured as small profiles as possible, the BIC index slightly favoured the solution with four profiles. The VLMR and LMR LRT indexes both favoured solutions with four ($p_{\text{VLMR}}, p_{\text{LMR LRT}} < .05$) and five ($p_{\text{VLMR}}, p_{\text{LMR LRT}} < .05$) profiles. Finally, the four-profile solution fit the interpretations made in previous studies (e.g., Parpala et al., 2010) and it contained a suitable smallest profile size.

Table 3. *Fit indices for profile solutions.*

	2 profiles	3 profiles	4 profiles	5 profiles	6 profiles
AIC	1370.584	1353.499	1337.643	1331.445	1319.864
BIC	1396.487	1390.503	1385.748	1390.653	1390.173
aBIC	1374.287	1358.789	1344.52	1339.91	1329.916
VLMR	-700.297	-678.292	-666.749	-655.821	-649.723
p_{VLMR}	0.0009	0.1424	0.0238	0.0094	0.1162
LMR LRT	41.579	21.81	20.648	11.523	16.61
$p_{\text{LMR LRT}}$	0.0013	0.1557	0.0279	0.0118	0.1324
Smallest profile (%)	14.72	5.35	5.02	0.33	1.67

The four profiles identified were named as follows (see the statistical characteristics of each profile in Table 8):

- Students applying a very deep approach (N = 116)
- Students applying a deep approach (N = 116)
- Students applying a dissonant approach (N = 52)
- Student applying a surface approach (N = 15)

The ANOVA comparison between the self-assessment groups (Table 4) revealed that differences between the profiles in terms of the study variables were rather drastic. There were significant differences regarding all of the variables of the study, with effect sizes varying from medium (achievement: .14) to extremely large (surface approach: .89). Tukey's post hoc testing showed that students in the deep approach profile reported higher levels of self-efficacy than those in the other profiles and outperformed them in terms of achievement. Because the surface approach profile was small (N = 15) and

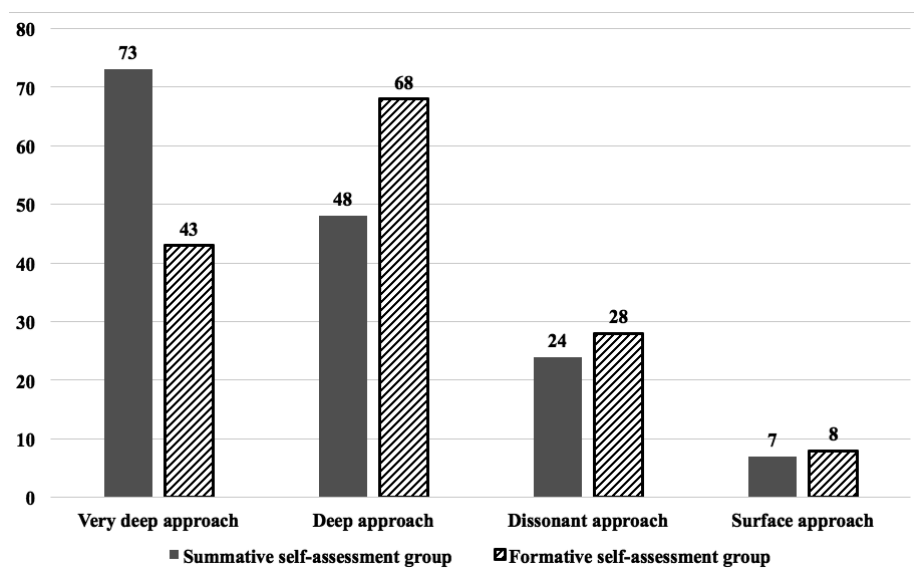
since the variance of self-efficacy was unequal in the student profiles, nonparametric testing was also conducted. The Kruskal–Wallis test further validated the significant differences between the student profiles regarding all the study variables ($p < .001$).

Table 4. ANOVA comparison between the student profiles.

	Profile 1 Deep approach (N = 116)		Profile 2 Low deep approach (N = 116)		Profile 3 Dissonant approach (N = 52)		Profile 4 Surface approach (N = 15)		ANOVA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F(4, 298)	η^2
Deep approach	4.07	0.67	3.57	0.69	3.34	0.78	3.12	0.79	20.01*	0.17
Surface approach	1.28	0.26	2.14	0.24	2.93	0.26	4.04	0.35	834.00*	0.89
Self-efficacy	4.60	0.46	3.96	0.62	3.41	0.87	2.79	0.88	68.13*	0.41
Achievement on the course	0.82	0.12	0.77	0.11	0.70	0.14	0.66	0.13	15.72*	0.14
* p < 0.001										
										Post hoc (Tukey HSD) comparison
										1 > 2,3,4
										4 > 1,2,3; 3 > 1,2; 2 > 1
										1 > 2,3,4; 2 > 3,4; 3 > 4

Comparisons were made within the student profiles in terms of the two self-assessment groups (for the distribution of the profiles see Figure 2). First, there were no significant differences in any of the profiles regarding deep and surface approaches. The profiles were consistent in terms of the course achievement as well. However, great differences with large effect sizes were found regarding students' self-efficacy, which was reported higher in the summative self-assessment group in the very deep ($M_{\text{summ}} = 4.74$, $SD_{\text{summ}} = .39$; $M_{\text{form}} = 4.37$, $SD_{\text{form}} = .48$; $t(115) = 4.46$, $p < .001$, $d = .83$) and deep approach profiles ($M_{\text{summ}} = 4.18$, $SD_{\text{summ}} = .50$; $M_{\text{form}} = 3.81$, $SD_{\text{form}} = .65$; $t(115) = 3.32$, $p < .001$, $d = .64$).

Figure 2 The distribution of the student profiles in the two self-assessment models.



3.4.4 DISCUSSION

Study I adds to the scarce literature on self-grading in higher educational settings, aiming to offer evidence for the scientific discussion about alternative summative assessment methods. Overall, the findings underline that students in both self-assessment groups reported a high level of deep approach and scored highly in terms of achievement; this is unusual in the context of natural sciences (Parpala et al., 2010). Study I continues the tradition of the literature on approaches to learning in seeking appropriate ways to support students' "deep shifts" (Haggis, 2003) in their studies. As inducing a deep approach has been shown to be tricky through assessment (Baeten, Dochy, & Struyven,

2008; Haggis, 2003; Struyven et al., 2006), it was interesting to find that summative self-assessment had been able to support a deep approach to learning better than formative self-assessment. Summative self-assessment was also strongly connected with higher levels of self-efficacy. Bandura (1997) suggested that students with strong self-efficacy set higher goals for themselves — it may be that the students in the summative self-assessment group were not only able to aim higher but also to aim differently, not for the teacher's sake but for their own.

The main implication of Study I was that if the self-assessment literature aims to move forward, it will not be enough to simply state that self-grading practices should not be used without offering any empirical evidence or investigation of the topic (e.g., Andrade, 2010; Andrade & Du, 2007; Bourke, 2018). If assessment were to be used seriously with the intention of supporting deep shifts, summative grading practices may have to be rethought. Study I encouraged researchers and practitioners alike to ask: How do summative grading methods offer opportunities for students to adopt a deep approach to learning? Study I also highlighted the importance of considering the interconnection between summative assessment and its various educational and political contexts. The kind of conceptual change in what is meant by self-assessment, as proposed by the summative self-assessment model, is not considered desirable or even possible in every context, even though as shown here, such a practice might have the potential to promote the quality of studying.

3.5 STUDY II

3.5.1 AIMS

Study II was the only one not to draw on the experimental research design as reported in Study I (and as examined in Studies III and IV as well). Instead, Study II was based on a summative self-assessment model implementation in a smaller undergraduate mathematics course. Through a mixed methods approach Study II examined student subgroups in terms of deep and surface approaches to learning and the various contextual factors students connected with these approaches. The study examined the student subgroups in terms of both course achievement and studying methods (approaches to learning).

3.5.2 METHODS

The data for Study II was collected using a questionnaire following an undergraduate mathematics course which utilised the summative self-assessment model. The course was the same linear algebra course as reported

in Study I, but was taught by another teacher. The course was offered at the Open University at the University of Helsinki during the summer; the 164 participants of the course reflected a heterogeneous group not only consisting of university students, as anyone could participate the courses at the Open University. 113 participants took part in the final survey and gave their active consent to take part in the study.

The data was collected through a survey after the students had followed the summative self-assessment model and graded their own course mark. The survey consisted of two parts. First, deep and surface approaches to learning were tested with a self-reported ETLQ-questionnaire as in Study I. Both scales consisted of four items ($\alpha = .62$ and $\alpha = .75$, respectively). The survey also included open questions in which the students could answer in their own words. These descriptive questions concerned students' experiences of the summative self-assessment model (e.g., "How did you experience the fact that there was no exam in this course?"), as well as the learning environment more generally ("How have you been able to benefit from the feedback during the course?"). The open questions loosely followed the interview protocol as constituted by Mumm and colleagues (2016). Data about course achievement was collected from the digital Moodle environment; this data included the scores of the mathematical tasks during the course (with automatic feedback, with student tutor feedback, and with no feedback) and the self-graded marks. As in Study I, students gave their active consent to participate in the study.

The analysis of Study II consisted of two parts. First, a hierarchical cluster analysis was conducted to examine the student subgroups in terms of deep and surface approaches to learning. The analysis utilised Ward's algorithm as the clustering algorithm, and the variables were standardised to Z-points before the analysis. The clusters were statistically compared and characterised through ANOVA comparison and further Bonferroni post hoc testing. The second part of the analysis consisted of a qualitative content analysis (Miles, Huberman, & Saldaña, 2014; Schreier, 2012) in which all the open question responses were used as the pool for analysis (rather than analysing these within each of the student clusters, for example). The conventional analysis (Hsieh & Shannon, 2005) sought expressions related to deep and surface approaches to learning. Overall, 74 analysis units consisting of a coherent, identifiable idea were identified and connected with the elements of the course design.

3.5.3 FINDINGS

The four cluster solution identified from the data is presented in Table 5. The number of clusters was fixed to four after observing the dendrogram of the data and after performing a discriminant function analysis that predicted 95,5 % of the cluster membership. Also, this cluster solution seemed appropriate to describe the data as it distinguished a small cluster consisting of students who

utilised a high level of surface approach to learning. It can be seen that this smallest cluster ($N = 10$) consists of students demonstrating a very high level of surface approach to learning and a considerate level of deep approach to learning. Also, the two biggest clusters (with both $N = 36$) consisted of students with a very high ($M = 4.38$) and high ($M = 4.06$) deep approach to learning.

Table 5. *The four cluster solution and scores regarding deep and surface approach to learning and course achievement*

		Tasks with					
		Surface approach	Deep approach	Automatic feedback (max. 64)	No feedback (max. 40)	Teacher feedback (max. 9)	Grade (scale 1-5)
Total	M	2.07	3.89	52.71	31.94	7.93	4.10
N = 113	SD	.70	.66	10.45	8.76	1.70	1.45
Cluster 1	M	1.90	3.14	51.16	30.10	7.57	3.75
N = 30	SD	.41	.51	10.68	8.19	1.96	1.74
Cluster 2	M	1.45	4.38	56.17	35.86	8.25	4.58
N = 36	SD	.28	.41	7.65	7.34	1.76	1.20
Cluster 3	M	2.40	4.06	51.76	30.92	7.94	3.97
N = 36	SD	.28	.41	10.75	8.56	1.57	1.45
Cluster 4	M	3.60	3.73	48.30	27.00	7.80	3.63
N = 10	SD	.39	.48	14.92	11.45	1.03	.92

Course achievement was compared between the clusters using ANOVA. The assumption regarding the homogeneity of variance was met for all variables (Levene test, $p = .06 - .73$). As cluster 4 only consisted of ten students, it was tested separately; all the variables of the study were normally distributed (Kolmogorov-Smirnov test, $p = .20 - .61$). Unsurprisingly, significant differences were found between the cluster in terms of surface approach ($df = 3$, $F = 128.9$, $p < .001$) and deep approach to learning ($df = 3$, $F = 45.19$, $p < .001$). Regarding the scores on course achievement, the only significant difference between the groups was found in terms of the scores from non-assessed tasks with no feedback ($df = 3$, $F = 312.27$, $p < .05$). Small yet insignificant differences were found between points from automatically assessed tasks ($df = 3$, $F = 243.68$, $p = .081$) and course grades ($df = 3$, $F = 4.73$, $p = .078$).

The Bonferroni post hoc test confirmed that the differences between the clusters were not extensive. All the clusters differed significantly in terms of surface and deep approaches ($p < .05$), except for clusters 3 and 4 that only differed in terms of their surface approach to learning. The students in clusters 1 and 4 were shown to have completed significantly fewer non-assessed

exercises with no feedback than the students in cluster 2 ($p < 0.05$). No other differences were found.

The findings of the qualitative content analysis showed that, in their responses, the most common factor that students connected with a deep approach to learning was a category named “Innovative assessment” (see Table 6). This category consisted of two main subcategories, titled Summative self-assessment and Various forms of feedback. Also, the students connected a deep approach with the course material and the learning objective matrix of the course, which were both described to be crucial for learning. In contrast, a surface approach to learning was most often identified from responses that were not concerned about this specific learning environment but about mathematical learning culture in general. Students linked “cramming” with traditional course exams. Many described how in other mathematics courses they might attempt to memorise the lecture notes a couple of days before the exam, yet this practice was not needed in this course. Also, some students linked the non-assessed tasks with surface learning by mentioning that they completed them with less effort than the rest of the tasks since no feedback was provided for them.

Table 6. *The contextual factors students connected with deep approach to learning in their open answers.*

Categories	Category subclasses	Category classes
No exam (22)	Replacing exam with self-assessment	Innovative assessment
Self-assessment tasks (13)		
Variety of feedback (4)	Various forms of feedback	
Formative feedback (3)		
Feedback from peers (3)		
Teacher-assessed tasks (1)		
Automatic feedback (2)	Course materials	
Learning objective matrix (6)		
Lecture notes (1)		Student-centred course materials

3.5.4 DISCUSSION

Study II painted a broad picture of the “end point” of one course adaptation of the summative self-assessment model: What were the achievement outcomes, and how had the students studied during the course? It was found out that the students largely reported high levels of a deep approach to learning, while only one small student cluster ($N = 10$) was identified in which the students utilised a high level of surface approach to learning. The four cluster solution showed

only minor differences between the clusters in terms of course achievement; this finding underlines the importance of not only focusing on learning results but also on the process leading up to them. The qualitative content analysis further strengthened the connection between summative self-assessment and a deep approach to learning.

3.6 A BRIEF SUMMARY

Overall, the quantitative Studies I and II painted a broad picture of how the summative self-assessment model relates to student subgroups, as investigated through deep and surface approaches to learning. Even though Studies I and II only consisted of a single survey study, both suggested that the summative self-assessment model was able to support a deep approach to learning in the context of undergraduate mathematics education, and in the context of large university courses in particular. While Study II hinted at the role of social and digital support systems in supporting a deep approach to learning, the question still remained: What was the reason for these findings? Furthermore, Study I did not offer an explanation for the difference in reported self-efficacy between the summative and formative self-assessment models. To elaborate on these findings, a further qualitative approach was needed. This idea initiated Studies III and IV, which will be introduced next.

4 BEYOND THE INDIVIDUAL: INVESTIGATING AGENCY AND POWER

Studies III and IV were conducted to elaborate on the findings of Studies I and II through student interviews. As Studies I and II drew on individual perspectives in their quantitative and mixed methods approaches, Studies III and IV supplemented these earlier studies through adopting a socio-cultural viewpoint. Through the socio-cultural approach, these studies framed the findings of Studies I and II in relation to their context of undergraduate mathematics education. This was conducted by using the theoretical frameworks of agency and power, that will be presented next.

4.1 STUDENT AGENCY

This doctoral thesis conceptualises student agency through a socio-cultural perspective rather than as a characteristic of the students themselves (Emirbayer & Mische, 1998). This choice echoes the call by Charteris and Smardon (2018) for researchers to consider socio-cultural and socio-material frameworks for agency in modern learning and assessment environments (see also Charteris & Thomas, 2017). The socio-cultural approach supplements earlier self-assessment research, where the phenomenon of students' control over their own learning is largely conceptualised through psychological perspectives in this field (see e.g., Panadero, Jönsson, & Botella, 2017; also Panadero, Jönsson, & Strijbos, 2016, for self-regulation). Through a socio-cultural approach, it is possible to examine the *affordances* that self-assessment offers for promoting student agency. Greeno (1994) defined affordances as properties of learning environments that contribute to enabling certain desirable actions. Using Kennewell's (2001) metaphor to describe the concept, self-assessment would ideally leave the door open for agentic learning — the choice to enter that door is left for the students themselves.

This doctoral thesis draws on the socio-cultural framework of *ecological agency* (Emirbayer & Mische, 1998) to conceptualise the notion of agency. Understanding agency from the ecological perspective makes it possible to “understand why an individual can achieve agency in one situation but not in another” (Biesta & Tedder, 2007, p. 137), building up the connection with the concept of an affordance (Kennewell, 2001). The ecological framework for agency underlines the social and cultural contexts and their interplay with affordances for agentic learning (Charteris & Smardon, 2018; Charteris & Thomas, 2017). Biesta and Tedder (2007) conclude that the ecological view sees the students acting by means of their environment, rather than simply in an environment (see also Biesta, Priestley, & Robinson, 2015). The task of

understanding the interaction between agency and self-assessment becomes a task to understand which self-assessment systems are “more conducive to developing the different modalities of agency” (Emirbayer and Mische 1998, 1005).

In their seminal work, Emirbayer and Mische (1998) consider the temporal nature of agency. The three temporal dimensions — the iterative, projective, and practical-evaluative dimensions — are tied to the context of this thesis: self-assessment in the context of undergraduate mathematics education. The three dimensions of ecological agency highlight that students’ perceptions of assessment are always tied to their past experiences, but also to the contexts of the present and the future. The *iterative dimension of agency* deals with how students’ agency is affected by their past experiences, in this thesis referring to earlier experiences of mathematics assessment. The *projective dimension* refers to “*the imaginative generation by actors of possible future trajectories of action, in which received structures of thought and action may be creatively reconfigured*” (Emirbayer & Mische, 1998, p. 971). Here, the projective dimension refers to students’ future intentions concerning self-assessment (Tan, 2007, 2009). Finally, agency can only ever be acted out in the present. The *practical-evaluative dimension of agency* is connected with the present enactment of agency, or, students’ experiences of self-assessment in the studied contexts of formative and summative self-assessment.

The notion of ecological agency avoids understanding agency as residing within individuals; rather, agency is enacted through students’ agentic orientations within their social and cultural contexts (Rajala & Kumpulainen, 2017). To conceptualise these agentic orientations in relation to self-assessment, this doctoral thesis utilises a division of agency into three different types. This choice follows the work of Harris and colleagues (2018) who argue that even though it is often assumed that agentic orientations during assessment practices would be directed towards academic growth, students might show their agency in *maladaptive ways*. This concern has been constantly raised in the self-assessment literature, as assessment dishonesty is seen as one of the main reasons for using self-assessment only in a formative way (e.g., Andrade, 2010; Andrade, 2007; Bourke, 2018). Therefore, in this doctoral thesis a distinction was made between *adaptive* and *maladaptive* agentic orientations. Furthermore, students might show a *lack of agency* during self-assessment if they were feeling helpless or unable to respond critically to these practices in that specific educational context.

4.2 POWER

No matter how many affordances an assessment practice offers for agentic learning, in the end those affordances are only manifested in practice through students’ intentions (or, through agentic orientations). The concept of *power*

is used in this doctoral thesis to critically examine the environmental factors that might hinder or promote students' use of those affordances. Following Kennewell's metaphor (2001) again: When the door to agentic learning is open, who are enabled to walk through it, and how?

This doctoral thesis continues the tradition of examining self-assessment in higher education through the lens of power (e.g., Leach, Neutze, & Zepke, 2001; Tan, 2007, 2008; Taras, 2001, 2008, 2016). In their widely cited article, Reynolds and Trehan state that assessment is the "primary location for power relations" in higher education (2000, p. 268). Often, the issues of power are connected with unilateral grading processes (Leach, Neutze, & Zepke, 2001). For example, Falchikov (2005) argues that traditional assessment "reinforces the power imbalance between teachers and learners" (p. 246). It has been claimed that these issues could be addressed through self-assessment (e.g., Taras, 2001, 2008, 2016). However, rarely has the self-assessment literature aimed at disrupting the already existing power relations of grading. In this doctoral thesis, the power relations concerning summative self-assessment are investigated based on the typology on the relationship between self-assessment and power by Kelvin Tan (2004). The typology sums up three theoretical frameworks for power, which are introduced here in detail: Sovereign power, epistemological power and disciplinary power.

Sovereign power. Sovereign power refers to historical power that is based on a unilateral relation between sovereign rulers and subjects (Patton, 2012; Tan, 2004). Earlier literature has applied the historical analysis on the dynamics of societies and civilisations to the field of self-assessment by focusing on the unilateral power relations that teachers wield over students (Reynolds & Trehan, 2000). Sovereign power manifests in episodic and interpersonal acts between the wielders of power and the subjects of it (Clegg, 1989). Since the manifestations of sovereign power can be identified rather simply, it has been characterised as a "straightforward conceptualisation" (Patton, 2012, p. 722) that oversimplifies the complex power relations of assessment (Taylor & Robinson, 2009). It is notable that in the episodic acts of sovereign power, there are always roles for actors based on unequal power relations. Since sovereign power can only reside in one individual actor at any given point of time, one must be either a ruler wielding power or the subject responding to it (Tan 2004). In assessment, the possible role for the teacher has traditionally been "the external assessor", while the students are the ones who are being assessed (Boud & Falchikov, 2006; Torrance, 2000). In the field of self-assessment, sovereign power can be identified when only one assessment outcome can be considered as valid: either the student's or the teacher's, since only one assessment result can be sovereign (Tan, 2004).

Epistemological power. The notion of epistemological power broadens the conceptualisation of sovereign power. Rather than focusing on interpersonal acts, epistemological power "affects teachers and students in the broader politics of institutions" (Tan, 2004, p. 654). Therefore, epistemological power is manifested only in relation to its contexts. Taras

(2016) uses the concept of *administrative engines* to illustrate how epistemological power acts behind both teachers' and students' actions and learning. Hanafin and colleagues (2007) argue that when a certain assessment practice is taken for granted in its own context, it becomes a part of the institutional epistemology; the notion of epistemological power identifies these processes.

Institutional epistemologies control what can be assessed as valid knowledge and how evidence about that knowledge can be collected (Tan, 2004). For example, in mathematics assessment in Finland the use of self-assessment is rare (Atjonen et al., 2019), which might imply that self-assessment does not fit the institutional epistemology of how mathematical knowledge can be assessed. As noted, these epistemologies can be identified from both students' and teachers' epistemological beliefs since both these groups take an active part in either maintaining or challenging these epistemologies (Hanafin et al., 2007; Tan, 2004). Moreover, students are aware of the institutional assessment culture which guides their behaviour during their self-assessment (Taras, 2008, 2016). From this perspective, the question of whether summative self-assessment empowers students is about whether the institutional epistemologies are challenged (Taras 2016). Furthermore, assessment practices are affected by larger political factors, making the analysis of epistemological power complicated. The choice of assessment practices is not done in a vacuum, but in a complex network of higher education policies, departmental regulations, and teachers' epistemological beliefs, for instance.

Disciplinary power. The third conceptualisation of power draws on the Foucauldian notion of disciplinary power. Disciplinary power is not only manifested in interpersonal acts or in institutional epistemologies. Rather, it is analysed by observing how power relations are produced through *discourses*. Foucault argues that discourses are “practices that systematically form the objects of which they speak” (Foucault, 1977, p. 49). Discourses produce meaning and knowledge and organise the ways we see the world (Foucault, 1982). Rather than being simply oppressive, disciplinary power is *productive*. Power imbalances are found when only certain actors have access to discursive processes and others are abridged as the subjects of disciplinary power (Foucault, 1977, 1982; Raaper, 2019; Patton, 2012). It is notable that discourses are not only restricted to language, but discursive practices are more broadly tied into communication and other ways of building knowledge.

Disciplinary power conceptualises assessment as a normalising technology of governmentality and control (Foucault, 1977, 1982). Tan (2004) warns that self-assessment might act as a mechanism for government of the self when students self-assess themselves only in relation to teacher-produced criteria (e.g., rubrics) which render them targets for surveillance: “Consequently, the students first learn to distrust their own judgements and then act as agents to constrain themselves” (p. 658). It has been claimed that the *self* in self-assessment is paradoxical as students' subjective identity is only constructed

in an institutional framework and then turned into a mechanism of control (Kasanen & Rätty, 2002), leading to students becoming police officers in their own self-assessment (Reynolds & Trehan, 2000). However, as Raaper (2019) reminds us, students are active actors in assessment situations and not just non-agentic subjects of disciplinary power. Students' beliefs and actions are constantly co-constructing the discourses around assessment, making their perspective an interesting starting point for research on disciplinary power.

4.3 RESEARCH OBJECTIVE

Studies III and IV reached beyond the individual and psychological premises of Studies I and II to understand summative self-assessment as a socio-cultural practice. As Studies III and IV were conducted as a part of the experimental design with the two self-assessment groups, they act as "sequels" for Study I; yet, the approach taken is substantially different. The overall research objective for these two studies was to examine students' perceptions of the affordances that summative self-assessment offers for the development of student agency, while acknowledging the complex power relations affecting the use of those affordances.

4.4 METHODOLOGICAL PREMISES

Studies III and IV drew on interview data that was collected as a part of the same research design as in Study I. After the course adaptation consisting of formative and summative self-assessment groups, an invitation to an interview was sent to all of the participants of the course via email. 41 interviews were conducted, of which 26 were with the students taking part in the summative self-assessment model and 15 with those who were taking part in the formative self-assessment model. The semi-structured interviews (24 - 62 minutes, 39 minutes on average) concerned students' experiences of self-assessment, and the interview protocol was the same for both self-assessment groups. The students were asked, for example, how they felt about assessing their own skills and whether they felt that they benefited from the self-assessment practices. The semi-structured interview protocol left room for the students' own personal experiences beyond the researcher-produced questions.

4.5 STUDY III

4.5.1 AIMS

In Study III, student interviews from both the summative and formative self-assessment groups were analysed from the viewpoint of student agency. Utilising the theoretical framework of ecological agency, students' agentic orientations were contrasted between the two self-assessment groups by examining how agency was indicated in students' accounts while reconstructing their self-assessment behavior. Study III sought to find out whether there were any specific features in agentic orientations in relation to the summative self-assessment model.

4.5.2 METHODS

The 41 students who participated in the interview study (26 from the summative self-assessment group, 15 from the formative group; referred to using the letters S and F, respectively, in the data excerpts) represented a rather heterogeneous group: There were 14 different majors represented, and the ages varied between 19 and 49. Only two of the students reported having previous experience of self-assessment in mathematics, reflecting the assessment culture of mathematics in Finland (Atjonen et al., 2019). As one of the students put it: "The earlier mathematics studies did not encourage me to self-assess".

The analytical process in Study III consisted of two parts, drawing on a data-driven and a theory-driven analysis method, respectively. The first analysis phase drew on in vivo coding (Saldaña, 2016) that used students' own words and sayings to let their voices be heard in the process. This first phase familiarised the researchers with the dataset and allowed them to reduce it to coded analysis units. These units were analysed in the second phase through a theory-driven qualitative content analysis (Schreier, 2012). The analysis utilised a 2 x 3 matrix with the temporal dimensions of agency (practical-evaluative and projective dimensions; the iterative dimension was omitted from the analysis since we expected the students in both self-assessment groups to have similar kinds of past experiences of mathematical self-assessment) and types of agency. The types of agency were coded as follows:

- maladaptive: accounts of agentially engaging with maladaptive behaviour such as cheating
- adaptive: accounts of agentially engaging with self-assessment to enhance learning
- lacking: accounts of not being able to critically respond to self-assessment

Finally, the findings were contrasted (Schreier, 2012) between the two self-assessment groups to identify differences in terms of agentic orientations. The contrasting process aimed to examine qualitative differences and nuances between these two data sources rather than to point out quantitative differences.

The data analysis was validated through a researcher triangulation (Denzin, 1978). The first author started by coding the first 20 interviews individually, after which the two authors discussed these findings and co-coded the unclear analysis units. The second author also conducted random checks for the data by recoding the data; this process sought to promote the internal validity of the first author’s analysis. After all the transcripts had been analysed, the same validation process followed again.

4.5.3 FINDINGS

Tables 7 and 8 present a descriptive overview of the analysed data, based on two approaches to quantify the results of the analysis.

Table 7. *The number of agentic orientation identified in the two self-assessment groups.*

	Practical-evaluative dimension			Projective dimension		
	Adapt.	Maladapt.	Lacking	Adapt.	Maladapt.	Lacking
Summative self-assessment group	111	11	80	26	5	10
Formative self-assessment group	35	15	31	9	7	8

Table 8. *The number and proportion of students who had at least one account of the certain agentic orientation.*

	Practical-evaluative dimension						Projective dimension					
	Adapt.		Maladapt.		Lacking		Adapt.		Maladapt.		Lacking	
	N	%	N	%	N	%	N	%	N	%	N	%
Summative self-assessment group	26	100	8	30.8	25	96.2	20	76.9	5	19.2	9	34.6
Formative self-assessment group	14	93.3	10	66.7	12	80.0	7	46.7	6	40.0	6	40.0

First, the analysis sought agentic orientations that were common to both self-assessment groups. In both groups, students largely reported adaptive-practical-evaluative agentic orientations that were connected with utilising self-assessment to promote students’ learning and studying during the course. Manifestations of this were, for example, the accounts of students being able to monitor their learning and studying through the self-assessment tasks:

S13: You could follow your own learning by realising that hey, I’ve completely ignored this thing over here. You could kind of control your own learning better.

F14: You become more conscious about the learning process because, well, just like in any sport, you constantly follow your own performance.

Maladaptive-practical-evaluative orientations mostly concerned stress and lack of time related to studying in both self-assessment groups. As the students had to negotiate how they would use their limited resources, self-assessment and self-reflection were described as something “extra” that could have been bargained amongst other study-related deadlines. Interestingly, as seen in tables 7 and 8, lacking-practical evaluative orientation was largely reported in both self-assessment groups. Self-assessment was described as a new, weird and complicated assessment method; students described this new kind of an assessment culture causing them to feel helpless, which was connected with the notion of lacking agency in the analysis. Also, transferring one’s new self-assessment skills, connected to the learning objective matrix, to real-life skills was described as being difficult for the students:

S15: This kind of self-assessment is based on previously-set learning objectives that can be linguistically defined into sentences. This requires an external expert who has already mastered the content. To me, it is extremely hard to try to learn something new on my own terms, so that I would set those clear goals for me, by myself. So, I wouldn't be able to produce a rubric. I couldn't assess my own skills with this kind of a mechanism.

Further analysis investigated those agentic characteristics that were specific to the summative self-assessment group. A frequent theme in these students' practical-evaluative accounts was "studying for myself". The students described that summative self-assessment enabled them to study linear algebra not for the examination but to gain personal knowledge that they could utilise in the future. An "extreme" version of lacking agency was also identified from the summative self-assessment group, as some students felt helpless about grading their own work — one student even refused to self-grade the course mark.

S12: Maybe my attitude wasn't that I would study in this course to be assessed. More like, my attitude was that I am here to build knowledge for myself through these tasks. It [self-assessment] changed my stance in terms of the assignments.

S21: If I'm studying for an exam, I often feel like now I'm studying for that exam. And for the fact that I would get a good grade. Now I felt more like I would have been learning to be able to use these skills in the future.

The most evident difference between the two self-assessment groups concerned the adaptive-projective agentic orientation. Only in the summative self-assessment group did the students connect self-assessment with future intentions to continue examining one's own learning through self-reflection. These students described changing their perceptions of assessment as they now understood self-assessment as a useful future skill.

4.5.4 DISCUSSION

Study III addressed various research gaps in the meagre field of assessment and agency in higher education. First, Study III confirmed the connection between formative self-assessment practices and student agency, as students in both self-assessment groups largely reported accounts of *adaptive-practical-evaluative orientation*. Self-assessment is often linked with concerns about students who cheat (e.g., Andrade & Du, 2007), but in this dataset it was the accounts of lacking agency that raised the authors' concerns. The students often reported not being able to monitor and evaluate their own

learning critically. This might reflect the examination-driven assessment culture of mathematics (Atjonen et al., 2019; Iannone & Simpson, 2011); self-assessment was largely reported to feel weird and even frightening, and students largely hoped for more support to promote their agency.

What characterised the summative self-assessment group were the students' accounts of future-driven adaptive-projective orientations, reflecting the earlier literature on future-driven self-assessment (Tan, 2007, 2008). Similar accounts were not identified from the formative self-assessment group, underlining the affordances for students' agency that only the summative self-assessment model was able to promote. Following Biesta and Tedder (2007), it was argued that this was because the *means* of self-assessment were different; self-assessment was done for the students themselves, not for the teacher in the summative self-assessment group. An "extreme" version of lacking agency was also identified from the summative self-assessment group, highlighting that an adequate support system to scaffold students' agentic studying is needed when agency is promoted through alternative assessment practices — especially in teacher-driven assessment environments.

Study III underlined the importance of understanding student agency through socio-cultural frameworks. Based on the findings, it is suggested that socio-cultural aspects must be considered if practical changes in assessment environments are promoted. Understanding the environmental affordances that different self-assessment models offer for promoting agency would supplement earlier psychological studies on self-assessment and self-regulation (Panadero, Brown, & Strijbos, 2016; Panadero, Jonsson, & Botella, 2017). For example, the findings of Study III showed that monitoring one's learning, which might fall under the category of "self-regulation", might actually reflect the *lack* of agency rather than adaptive forms of it. Study III calls for innovative assessment research to tackle the issues relating to the hindering of agency through assessment. Here, it was shown that summative self-assessment offered an affordance for future-driven agency, which could not be identified in terms of formative self-assessment. To sum up, in Study III it was implied that if promoting agency is truly to be the key feature of new generation assessment environments (cf. Charteris & Smardon, 2018), the summative assessment methods should be carefully reconsidered.

4.6 STUDY IV

4.6.1 AIMS

Study IV examined the same 26 interviews with students taking part in summative self-assessment that were analysed for Study III, yet this time focusing on the notion of power. Study IV continued the tradition of tying self-

assessment with the attempt to empower students, yet utilised a critical perspective that sought to understand whether “empowerment” actually happened — and, furthermore, theorised that concept of empowerment. By drawing on the theoretical frameworks of sovereign, epistemological, and disciplinary power, students’ conceptions of summative self-assessment were examined. The purpose was to understand whether the usual power relations of undergraduate mathematics assessment were disrupted through the assessment model, as reflected in students’ interviews.

4.6.2 METHODS

The participants of Study IV consisted of those 26 students who took part in the summative self-assessment model in the experimental study (Studies I, III, IV). The students formed a heterogeneous group, as they represented nine different majors and their ages varied between 19 and 45. Fifteen students were in their first year of study at university, and six students had studied for over five years.

The data analysis consisted of two phases, drawing on data-driven and theory-driven methods, respectively. First, the data was analysed using thematic analysis (DeSantis, & Ugarriza, 2000; Saldaña, 2016). The transcripts were divided into analysis units based on the themes that were identified from the data by utilising *in vivo* coding (Saldaña, 2016) that used the words and sayings of the students as codes. These coded units were categorised by connecting the themes into 26 meta-themes that shared the same thematic elements. The second analysis phase drew on theory-based elaborative coding (Auerbach & Silverstein, 2003; Saldaña, 2016). This phase was conducted to develop previous theoretical frameworks in the context of summative self-assessment. Each of the three notions of power were used as theoretical lenses through which to conceptualise the data-driven themes and meta-themes in order to understand the students’ conceptions. This was achieved by re-coding and re-grouping the themes and meta-themes of the first analysis phase. This phase was repeated until the regrouped themes produced coherent categories and a comprehensive understanding of each of the notions of power.

4.6.3 FINDINGS

Students’ conceptions reflecting sovereign power were connected to the idea of whose self-assessment was sovereign; or, whether power was shared during the process of summative self-assessment. Empowerment was identified in those accounts in which students described the feeling of sovereignty in the grading process and were able to share their resources in a new kind of a way. For example, the students largely reported that they were now able to take

control of their own studying. Agentic studying needed to be supported. For instance, one student described feeling empowered after utilising the course rubric to overcome their own self-criticism based on their clinical depression. However, not all of the students described feelings of empowerment — some thought that sovereignty was never shared. This was seen when students questioned the validity of summative self-assessment. Also, one student refused to grade themselves as they felt that they were not qualified to do that. Interestingly, some students connected the digital feedback offered in the digital self-assessment tasks with sovereignty, as they pointed out the “true” nature of one’s learning.

The notion of epistemological empowerment was identified in those accounts that questioned summative self-assessment as a natural part of mathematics assessment; in other words, the boundaries of mathematics assessment were pushed (see Leach, Neutze, & Zepke, 2001). Empowerment in terms of epistemological power was identified when students contested the assessment culture of mathematics and showed resistance towards it. In other words, some students were able to challenge the institutional epistemologies of mathematics assessment and described changing their studying methods after the course. Epistemological empowerment was also manifested as the students pondered the validity of the summative self-assessment model. According to some students, self-assessment was “the only way to know whether someone has actually learned”. Many students compared the validity of the model with course examinations. Disempowerment manifested in those accounts in which students hoped for an examination that would have “ensured” that they had learnt the mathematical content. Overall, the analysis of epistemological power revealed how deeply the institutional epistemologies of mathematics assessment were reflected in the student interviews; it was evident that the students had a clear understanding that mathematics has to be assessed through examinations, and they had adopted their role in this process long before their studies at university. Summative self-assessment did not fit these epistemologies; as one student noted, mathematical knowledge has to be “measured in a brutal way”. Many students were incoherent in their accounts, as they tried to make use of summative self-assessment but were only able to do that within the already existing epistemologies.

The analysis for disciplinary power understood summative assessment through a Foucauldian lens. The notion of empowerment was identified as students described being able to see themselves as critical and reflective actors rather than the recipients of assessment. Indeed, “studying for myself” was a frequent theme in the data. This reflective thinking changed the nature of learning during the course. For example, the feedback offered within the learning environment did not just represent “the truth” but acted as the base for students’ own active reflection. One student summarised this idea by claiming that it is important to “critically examine your own understanding yourself rather than let someone external do that”. The students often thought that someone was watching over their self-assessment, reflecting earlier

concerns of self-assessment as a technology of self (Foucault, 1982). This *someone* was often the teacher or the examination — or the students themselves. The students largely discussed “collecting evidence” and “proving” one’s learning. Furthermore, the self-surveillance was identified as the students wondered how to self-assess correctly. This was seen in the accounts that portrayed the process of summative self-assessment as a narrow pipe: the rubric with its exemplars told the students what to learn and how to self-assess, and the feedback system checked whether they had completed this process properly.

4.6.4 DISCUSSION

Through the three frameworks of power as utilised in Study IV, summative self-assessment was conceptualised as a socio-cultural practice. As the grading procedures of higher education (Boud, 2007; Torrance, 2007) and mathematics education (Study III) have even been claimed to hinder learning, it was argued in Study IV that there is a need to reimagine assessment practices in this context. In Study IV an attempt to disrupt power relations of undergraduate mathematics assessment was reported; the study sought to make sense of the complex power relations surrounding “assessment utopia” of summative self-assessment (see Filene, 1969). The notion of sovereign power, even though sometimes labelled as an oversimplification (Patton, 2012; Taylor & Robinson, 2009), enabled an interesting perspective into how students negotiated the sovereignty in assessment — in a situation that arguably offered *them* the sovereignty. The analysis of epistemological power considered the context of mathematics and the institutional epistemologies related to it (Hanafin et al., 2007). The results underline that asking students to challenge the existing epistemologies is not a neutral act. Finally, the Foucauldian notion of disciplinary power shed light on how summative self-assessment enabled both reflective learning (“empowerment”) and further governing through the governing of the self (“disempowerment”).

Study IV investigated the complexity of disrupting the power relations of assessment from within the broader system. Similar concerns have been reported in higher education studies for decades (see Filene, 1969). Study IV reminded us that researchers have an ethical responsibility to show resistance towards teacher-driven assessment and grading practices to “empower” students — and at the same time, it sheds light on the complex power relations that tie not just the pedagogical practices but the mere act of empowering students (Leach, Neutze, & Zepke, 2001). Just as Filene’s experiment of over 50 years ago, that depicted in Study IV was far from perfect. However, in Study IV it is argued that summative self-assessment did disrupt the power relations of undergraduate mathematics assessment as examined through all the three notions of power.

4.7 A BRIEF SUMMARY

Studies III and IV deepened the findings of Studies I and II through a socio-cultural approach. Rather than seeing summative self-assessment as a neutral practice that affects students' individual learning and studying in various ways, Studies III and IV investigated summative self-assessment itself as a socio-cultural practice. After examining summative self-assessment and its interplay with the quality of learning, student agency, and power, the questions arose: How to synthesise these findings? As the approaches in Studies I-IV drew on differing and even conflicting epistemological premises, a simple summary of the findings would not do justice for their divergence. It became evident that the four studies should be synthesised in a way that would not only build bridges between various theories but highlight the important ruptures between the individual and socio-cultural approaches. Next, based on these four studies, such a synthesis is presented, aiming to further theorise the interplay of agency and power.

5 SYNTHESIS THROUGH A DISCURSIVE-DECONSTRUCTIVE READING

After introducing Studies I-IV, this doctoral thesis synthesises these past works by taking a discursive turn. Studies I-II aimed to understand whether, and how, summative self-assessment could support students' learning and studying, and Studies III-IV further elaborated on these findings through a socio-cultural approach on agency and power. Finally, in this section, all these four studies are reinterpreted and synthesised drawing on the notion of *discourse* by Michel Foucault. Through his discursive concept of *subject positioning*, the concepts of agency and power are tied together to further theorise their interplay in relation to summative self-assessment. Thus, the synthesis addresses the positions that summative self-assessment enabled for students to take, and the agency within these positions. As earlier research has concerned the non-agentic position of "the assessee" that teacher-driven assessment might construct for students in higher education, this synthesis revisits Studies I-IV to examine whether this position was disrupted through summative self-assessment.

The discursive approach as taken in the synthesis brings forth the socially constructed nature of reality. This approach was chosen as it highlights both the important connections and ruptures between the varying methodologies and epistemological premises of Studies I-IV. More precisely, this synthesis builds on discursive-deconstructive reading (Ikävalko & Brunila, 2019) of the four studies of this doctoral thesis; reinterpreting Studies I-IV through such deconstructive approach sheds light on how students were positioned through the summative self-assessment model, but also on how they were positioned in the research concerning such assessment. This synthesis consists of four sections. First, an overview and a summary of Studies I-IV is presented through the tradition of triangulation. Next, the theoretical framework of subject positioning is presented, followed by the introduction of the discursive-deconstructive reading as the form of synthesis. Finally, the findings of such deconstructive synthesis are reported, and based on the synthesis, the concept of transformative self-assessment is constituted.

5.1 SUMMARY OF THE METHODOLOGICAL APPROACHES

To sum up the methodological approaches of this doctoral thesis, the classic typology of triangulation methods by Denzin (1978; see also Miles & Huberman, 1984) is employed. Based on this typology, triangulation is divided into *data triangulation* (utilisation of multiple data sources such as

interviews, surveys and documents), *researcher triangulation* (collaboration of many researchers in the same research project) and *method triangulation* (utilisation of many methods to gather and analyse data). Traditionally, triangulation has been advocated to verify one's findings through multiple perspectives (Meijer, Verloop, & Beijaard, 2002; Miles, Huberman, Saldaña, 2014). As Meijer and colleagues (2002) note, it is important in social sciences to gather a comprehensive understanding of the studied multifaceted phenomena. This doctoral thesis draws on all of the triangulation methods above and takes part in the triangulation of triangulation methods, or *multiple triangulation* as Denzin (1978) put it. For example, the four studies of this thesis are based on multiple data sources, and the data collection and data analysis methods utilised in each of them vary. To highlight the importance of each of these studies as individual perspectives aiming to understand the phenomenon of summative self-assessment, each of these methodologies are summed up in Table 9.

Methodological triangulation is largely utilised in this doctoral thesis. First, Studies I and II drew on profiling methods, with Study I using latent profile analysis and Study II cluster analysis as its main research method. The aim of these profiling analyses was to build a larger picture of what kinds of studying profiles could be identified after the courses. Studies III and IV took a different perspective as they drew on qualitative analyses. Both studies were based on a dataset consisting of student interviews. Studies III and IV utilised a data-driven approach to first divide the dataset into smaller analysis units (themes and meta-themes), followed by a theory-driven analysis (qualitative content analysis in Study III and elaborative coding in Study IV). Both studies used data as the basis for developing theory, and the interaction between theory and practice in particular.

This synthesis moves further from those triangulation strategies often connected with qualitative research that aim to foster “trustworthiness”, as Miles and colleagues put it (2014, p. 262). This synthesis specifically focuses on theoretical triangulation (Denzin, 1978; Van Drie & Dekker, 2013) to increase understanding of the interplay of power and agency in relation to self-assessment. According to Dekker (1978, p. 307), theoretical triangulation is most appropriate in “areas characterised by high theoretical incoherence”. The theoretical triangulation in this thesis does not only draw on mixing up theories, but rather perspectives that build on different epistemological premises (Table 9). All of these unique perspectives build up a comprehensive understanding of the mechanisms of power and agency in the context of university mathematics. Next, the concept of subject positioning is introduced, as it is utilised to synthesise these approaches to reframe the interplay of agency and power in terms of summative self-assessment.

Table 9. Overview of the most important triangulation methods utilised in the thesis.

Study	Ontological premises	Methodological approach	Data source	Analysis method
I	Positivist	Quantitative	Survey study (N = 299)	Latent profile analysis
II	Positivist	Mixed methods	Survey study with open answers (N = 113)	Cluster analysis, qualitative content analysis
III	Socio-cultural	Qualitative	Interviews (N = 41)	Thematic analysis, qualitative content analysis
IV	Socio-cultural	Qualitative	Interviews (N = 26)	Thematic analysis, elaborative coding

5.2 THEORETICAL FRAMEWORK: POWER AND SUBJECT POSITIONING

This synthesis further theorises the interplay of agency and power in relation to summative self-assessment through the concept of subject positioning, which draws on the idea of discourses as introduced by Michel Foucault. While Foucauldian understanding of discourses offers various theoretical tools to conceptualise assessment, this doctoral thesis focuses on the notion of subjectifying in particular. Discourses and disciplinary power produce knowledge but also *subjects* (e.g., Foucault, 1977, 1982; Heller, 1996; Phillips, 2006). A Foucauldian investigation of power focuses on understanding how people are objectified into subjects. As Foucault noted: “Discipline ‘makes’ individuals; it is the specific technique of power that regards individuals both as objects and as instruments of its exercise” (1977, p. 170). Subjects are not roles or identities, but rather fluid states that are constituted through discursive practices (Arribas-Ayllon & Walkerdine, 2008). The subjects are formed in the complex network of power relations under governmentality.

This means that subjects do not wield any kind of power themselves, but are passive objects of governing (Heller, 1996). Indeed, governmentality includes processes that shape and control its subjects (Bagger, Bjorklund Boistrup, & Norén, 2018); this doctoral thesis focuses on the processes of governmentality that assessment holds in undergraduate mathematics education.

Subjectification is executed through technologies of power (Foucault, 1991; Bagger, Bjorklund Boistrup, & Norén, 2018). While various actors take part in subjectifying discursive practices, this study concerning student self-assessment especially focuses on *technologies of the self* (Foucault, 1988). Foucault's later work (Foucault, 1990) concentrated on how governmentality works not only through external governing but through the government of the self. Foucault elaborated on technologies of the self, highlighting their role in the government of the self:

Technologies of the self, which permit individuals to effect by their own means or with the help of others a certain number of operations on their own bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality. (Foucault 1988, p. 18)

Power constructs a version of the reality in which we position ourselves as subjects - as particular types of people. Here, the Foucauldian notion of *subject positioning* is utilised (Foucault, 1977, 1982, 1988). Subjects occupy rather stable - yet discursive - positions within discourse, and these positions control what can be done, said and thought within a certain socio-cultural context (Arribas-Ayllon & Walkerdine, 2008). Subjects can only be constituted within these subject positions. As Phillips (2006) noted, subjects are tied in their social existence through their positions, and these positions "provide a recognised and recognisable form of subjectivity appropriate to that subject position" (p. 314). Subject positions are maintained through governmentality and the government of the self (technologies of the self). Often, the subject positions are made to look natural, as will soon be proven in terms of mathematics assessment. Subjects can only negotiate their positions within the power relations governing those positions in the first place. Analysing subject positioning focuses on identifying the "cultural repertoire of discourses available to speakers" (Arribas-Ayllon & Walkerdine, 2008, p. 118).

Within subject positions, possibilities for *agency* are constructed. Through this agency, subjects can reflect on their positioning and show resistance; as subject positions are socially constructed, the power dynamics related to them can always be renegotiated (Brunila & Ikävalko, 2012; Burr, 2015; Ikävalko & Brunila, 2017). As Burr (2015, p. 212) notes, the notion of subject positioning "allows us a conception of agency that acknowledges both the constructive force of discourse at a societal level as well as the capacity of the person to take up positions for their own purposes". It should be noted that agency as defined in relation to subject positioning differs from the ecological conceptualisation as defined in Study III. Agency, in discursive terms, refers to students'

capability to renegotiate their positioning through self-reflective discourses (Arribas-Ayllon, 2008).

5.2.1 SUBJECT POSITIONS IN MATHEMATICS ASSESSMENT: SELF-ASSESSMENT AS RESISTANCE

This doctoral thesis aims to understand the governmentality of undergraduate mathematics assessment through the investigation of subject positioning. According to Alhanen (2018), governing works by making its subjects visible through measurement, classification and evaluation of knowledge (see also Foucault, 1977; Saari, 2011; Torrance, 2000). Therefore, it is not surprising that assessment has been identified as a societal arena for power relations and governmentality. As Saari (2011) argues, assessment standardises its subjects through measurement and renders them compatible. Hanson (2000) stated that assessment *constructs* reality rather than reveals it through measurement, causing assessment to be the purpose for education. Torrance (2000) has elaborated on the discursive dimension of assessment especially in relation to examinations. Examinations organise and legitimate knowledge in testable form, which as an act already imposes that knowledge can be measured. According to Torrance, examinations are “perfectly appropriate for people performing certain roles to organize and test knowledge” (p. 177), yet warned that examinations are socio-cultural artefacts rather than natural ones. The discursive approach reminds us that examinations are not “bad” but naturalised, and that both teachers and students voluntarily submit to them. Torrance (2000) argues that the roles in the assessment process produce a hierarchy, and elaborates:

The very fact that we allow ourselves to be subjected to examinations, or subject others to them, validates and endorses the construction of identity through discourses of ‘passing and failing’, ‘knowing and not knowing’, defining who becomes one sort of person and who another. (2000, p. 178)

While acknowledging the importance of broader societal overviews (e.g., Saari, 2011), this doctoral thesis operates in its Finnish context of undergraduate mathematics education. In higher education, there has been a concern about the wider-level nature of assessment that only enables students the position of the assessee. For example, Boud and Falchikov (2006) claim that students would be seen as the subjects of assessment rather than as active agents. Evans (2011) analysed higher educational assessment documents in Australia, identifying the subject position of “student as a performer”. Evans claims that this discourse does not emphasise learning, but it emphasises awards (such as grades) that are accessed through performance: “this makes the student more

like an actor on a stage, performing and achieving an award for their performance, rather than someone interested in learning” (p. 221).

In undergraduate mathematics, performance is most often conducted through examinations (see Iannone & Simpson, 2011 for a British review). The notion that students prefer traditional assessment in undergraduate mathematics (Iannone & Simpson, 2015a, 2015b) might imply that they are acting from the position of the performer whose actions are assessed by someone else. This concern seems appropriate in the Finnish context, as in the lower levels of education mathematics assessment is mainly based on teacher-driven practices such as examinations, and alternative practices such as self-assessment are rarely introduced (Atjonen et al., 2019).

This doctoral thesis does not only focus on examining the positions in mathematics assessment but also aims to disrupt them. Thus, mathematical self-assessment is conceptualised through the concept of resistance. Self-assessment then becomes a political act; or, “a breach of self-evidence, of those self-evidences on which our knowledges, acquiescences and practices rest to show that things ‘weren’t as necessary as all that” (Foucault, 1991, p. 76). Lilja and Vinthagen (2014) argue that resistance towards disciplinary power and governing can be shown through the same technologies of self that form the basis of the government of the self (Allen, 2011; Phillips, 2006); following Foucault (1977) they state that power produces the potential of the subject to position itself differently. Resistance towards governing, and the government of the self in particular, is based on a subject’s critical reflection towards one’s subject positioning: “agency is contingently made possible by a kind of self-reflectivity” (Lilja & Vinthagen, 2014, p. 111).

Allen (2011) connects resistance towards disciplinary power with the notion of autonomy, underlining that autonomy does not only consist of critical reflection but of engaging in practices of self-transformation as well. This notion reframes mathematics assessment that would promote resistance, as this kind of assessment should not only foster critical thinking but offer concrete tools for self-transformation towards reflection and renegotiation of one’s subject positionings. However, resistance is only constituted within the governing it aims to disrupt, which has to be considered. As Allen noted, “since there is no outside to power, freedom always involves strategically reworking the power relations to which we are subjected” (2011, p. 51).

This doctoral thesis does not aim to deepen the theoretical understanding of the conceptual connections between governmentality, subjectification and subject positioning of students, but applies all these terms to understand the interplay of agency and power in self-assessment. As has been noted (e.g., Heller, 1996; Foucault, 1988; Phillip, 2006) subjects can only operate through their subject positions. Subject positions control agency for the subject tied to that specific position in its resources for repositioning itself. Also, as shown, resistance in mathematics assessment would not only foster critical thinking but offer tools for self-transformation by utilising the technologies of self that

are the basis for governmentality of assessment as well (Allen, 2011; Foucault, 1977; Lilja & Vinthagen, 2014).

Finally, as this doctoral study investigates student positioning as constructed in Studies I-IV, the findings of the synthesis also reflect the researcher's positioning (cf. Brunila & Valero, 2018). As a researcher, I have only worked through my own researcher positioning while conducting these studies. Through this researcher position, students taking part in this doctoral study have been positioned as certain kinds of participants, which will be discussed later on.

5.3 DISCURSIVE-DECONSTRUCTIVE READING: JOINING FOUCAULT'S GAME TO UNDERSTAND SELF-ASSESSMENT

The synthesis of this doctoral thesis draws on discursive-deconstructive reading of Studies I-IV (Brunila and Ikävalko 2012; Ikävalko and Brunila 2019; Naskali, 2003). Through this approach, this thesis makes sense of the four studies and investigated subject positioning of the students in them. Discursive-deconstructive reading concerns "both discursive power relations and the functioning in them, in other words, how the subject is formed" (Ikävalko and Brunila 2019, 3). Studies I-IV are used as discourse samples (Arribas-Ayllon & Walkerdine, 2008; Fairclough, 1992) to examine and contest the positions that were enabled to the students through the summative self-assessment model - and through the research conducted in relation to it. Discursive-deconstructive reading is not exactly an analytical method but an "analytical tool that can expose and dissect artificial oppositions" (Ikävalko & Brunila, 2019, p. 35) such as student/teacher and assessee/assessor. Ikävalko and Brunila (2019) note that discursive-deconstructive reading is suitable for analysing the interaction of agency and power in particular. The discursive approach draws on Foucault's notions of discourses and discursive practices, that have been defined as "practices of knowledge formation" that focus on "how specific knowledges ('discourses) operate and the work they do" by Bacchi and Bonham (2014, p. 174; see also Alhanen, 2018). This doctoral thesis examines those discursive practices that position students in Studies I-IV.

This doctoral thesis does not aim to participate in the theoretical-philosophical - and decades-long - discussion about the role and formation of discourses, nor does it aim to develop Foucault's theoretical ideas further. Instead, following Ikävalko and Brunila (2019), the intention of this doctoral thesis is to apply these ideas in practice, in a specific context; not only to develop theory but to evoke practical change (see also Brunila & Ikävalko, 2012). This perspective reflects Foucault's ideas as well, as he famously invited everyone to apply his theories in practice and join the game rather than treating the theories as dogmatic truths; after all, he claims his own work to be

“philosophical fragments put to work in a historical field of problems” (Foucault, 1991, p. 74). As will be argued, the deconstructive approach positions this doctoral thesis as a political text, rather than as an apolitical one.

5.4 THE OBJECTIVE AND PROCESS OF THE SYNTHESIS

Earlier studies have highlighted the non-agentic position that teacher-driven assessment practices construct for students through governmentality (e.g., Torrance, 2000); the objective of the deconstruction process of student positioning aimed to examine whether, and how, the “natural” position of the assessee, the one being assessed, was contested through summative self-assessment. The goal of the deconstruction process was to further develop the conceptual understanding of the interplay of student agency and power in assessment in the context of undergraduate mathematics education. Yet, as Studies I-IV all drew on concrete course implementations of summative self-assessment, this synthesis also reports a practical attempt to disrupt the non-agentic positioning processes as connected to assessment in higher education.

Given its multidisciplinary nature, the deconstructive synthesis of this doctoral thesis aims to contribute within various scientific fields. First, this synthesis widens the theoretical understanding on self-grading practices in higher education. Moreover, this synthesis builds on earlier studies concerning self-assessment and power (e.g., Tan, 2004, 2007) by reframing the disciplinary power relations through the notion of subject positioning. Furthermore, this synthesis develops understanding of self-assessment as a socio-cultural practice. Finally, as earlier work on self-grading in higher education has been conducted mainly with small class sizes (e.g., Milne, 2009), this synthesis reports an attempt to disrupt the “natural” positions of assessment in the context of large courses.

The deconstruction process identified the “cultural repertoires” of discourses (Arribas-Ayllon & Walkerdine, 2008, p. 118) of students across the datasets of Studies I-IV. This was achieved by rereading each of the research articles, and their findings in particular. The analysis sought to understand the ways that students both contested and further maintained the position of “the assessee”, or the receiver of assessment. The deconstruction process investigated whether the students were enabled to agentially renegotiate their positioning in the process of summative self-assessment. Finally, student positioning in relation to the research designs is deconstructed to highlight the ruptures between Studies I-II and Studies III-IV; these studies already positioned students as certain kinds of subjects through their epistemological and methodological approaches. The deconstruction process of students as participants consisted of reinterpretation of the methodology sections of Studies I-IV.

5.5 THE DECONSTRUCTION OF STUDENT POSITIONING IN STUDIES I-IV

5.5.1 DISRUPTING THE POSITION OF THE ASSESSEE

What was identified from students' discourse in Studies III and IV was a disruption of the subject position of "the assessee", which will be described in this section. The deconstructive reading identified the cultural repertoires of discourses enabled for the students taking part in the summative self-assessment to disrupt this position. First, the students taking part in summative self-assessment showed accounts of technologies of self that were not utilised for government of the self but for reflective self-subjectifying (Studies III and IV). "Finally studying for myself, not for the exam" was a frequent theme across the interview dataset that manifested the technologies of self. As reflective technologies of self were now part of students' discursive repertoire, the students were now enabled to renegotiate their positioning as "the assessee". The reflective space that was opened up by summative self-assessment further enabled students to critically contest the "natural" assessment practices of mathematics by conceptualising them as socio-cultural artefacts rather than as given truths. Therefore, summative self-assessment not only made the students more aware of reality, it *constructed* reality (cf. Hanson, 2000). Raaper (2019) has suggested that students are active negotiators of power relations of assessment; this doctoral thesis underlines the role of subject positioning in the process. Studies I and II were interpreted as indicators for reflective negotiation of student positioning happening on a broader level; this perspective will also be contested in Section 5.5.4.

Study I identified the projective agentic orientation (Emirbayer & Mische, 1998) of the interviews of students in the summative self-assessment group, reflecting the notion of future-driven self-assessment by Tan (2007). This future-driven discourse enabled the students to conceptualise their learning as a part of their future-driven skills and abilities, and not only in terms of the linear algebra course at hand. Moreover, as one student put it: "My attitude was that I am here to build knowledge for myself through these tasks." The reflective discursive practices enabled for students through sovereignty in self-assessment reframed the nature of learning mathematics; this finding challenges the governing of assessment (cf. Torrance, 2000). The mathematical skills and competences were now framed as something that the students were learning to build their own personal knowledge for the future. This process was largely interpreted to disrupt the positioning of students as assessees.

While the position of the assessee was contested and disrupted, it was evident that one single summative self-assessment implementation was not sufficient to completely reframe student positioning. This was seen in the

incoherent student perceptions about self-assessment in Studies III and IV; one single student might have been able to contest their own positioning as the receiver of assessment, while at the same time hoping for a final examination to validate learning. Furthermore, reflectivity over the position of the assessee did not necessarily manifest as desired student behaviour, such as “promoted deep approach to learning”. Rather, just as Harris and colleagues noted (2018), enabling agency to students might manifest in maladaptive ways. An example of such maladaptive behavior is an interview reported in Study III, in which a student recounted how they were now able to observe their own learning processes reflectively; yet, the student chose to focus their resources on other mathematics courses and family issues rather than on learning linear algebra.

5.5.2 THE DISRUPTIVE PRACTICES THAT ENABLED AGENTIC POSITIONING

As Allen (2011) noted, resistance against disciplinary power builds not only on self-reflection, but also on concrete tools for self-transformation. Following her work, the disruptive practices were identified from the qualitative data of Studies II, III and IV to shed light on those elements of the summative self-assessment that opened up the reflective space for students to renegotiate their positioning. In this section, the main disruptive practices are introduced: *sovereignty*, *the lack of an examination* and various *support mechanisms*. It should be noted that all of these three practices were utilised together, and that their potential for disruption is only constructed within the socio-cultural contexts of this study; the same practices might not be considered disruptive in other contexts beyond undergraduate mathematics education.

Study IV underlined the importance of *sovereignty* while employing self-assessment in an empowering way. Even though the notion of sovereign power has been claimed to oversimplify the complex interconnections between assessment and power (Patton, 2012; Taylor & Robinson, 2009), in Study IV it was suggested that it was indeed sovereignty that opened up the reflective space for students to reflect on their positioning. The students largely reported taking control of their own studying, which is indicated as a renegotiation of one’s positioning as “the assessee”. This finding is interpreted through the notion of technologies of self, as the shift in sovereign power made the “natural” subject positioning of the assessee visible to the students.

Disrupting the sovereign power relations through summative self-assessment excluded the course examination as the sovereign grading method. Indeed, Studies II and IV underlined that *the lack of an examination* itself was connected with students’ agentic subject positioning. In students’ responses in Study II, the category of “no examination” was identified more often than the category of “the formative self-assessment tasks”, for instance. One student commented in Study IV: “If I’m studying for an exam, I often feel like now I’m

studying for that exam. And for the fact that I would get a good grade. Now I felt more like I would have been learning to be able to use these skills in the future.” These findings are interpreted through the notion of governmentality, as examinations have often been connected with the governing of assessment (cf. Reynolds & Trehan, 2000; Saari, 2011). It might be that in contexts where traditional assessment methods dominate, such as the context of undergraduate mathematics education, removing the examination is sufficient to open up the reflective space for students’ agentic positioning through the technologies of the self. As Study IV suggested, the lack of examinations acts as a disruptive practice when it challenges the institutional epistemologies (cf. Hanafin et al., 2007); in the context of undergraduate mathematics education, simply removing the course examination pushed the boundaries of the institutional epistemologies of this specific context.

Removing the course examination and replacing that with self-grading would not have been sufficient in itself to tackle the issues of disciplinary power that earlier research has connected with self-assessment (e.g., Kasanen & Rätty, 2000; Tan, 2004). Instead, many elements of the summative self-assessment model were needed for supporting students’ reflective negotiation of their positioning. For example, the role of the rubric was raised in Studies II, III and IV as it enabled students to reflect on their own learning and set goals for themselves. The rubric acted as the main resources for practicing self-assessment formatively; other resources included, for example, the formative self-assessment tasks themselves and the feedback offered from those tasks. Framing summative self-assessment as a process (López-Pastor et al., 2012) supported students’ re-negotiation of their positioning through technologies of self. Self-assessment was practiced, and during the process students were asked to engage with feedback offered by student tutors, the digital system and their peers.

These findings highlight that various concrete support mechanisms are needed when the position of the assessee is disrupted within a teacher-driven assessment culture. Without those mechanisms, practices meant to disrupt might fall flat - or even further strengthen the position of the assessee. This finding underlines the importance of conceptualising the support mechanisms for agentic positioning themselves as socio-cultural practices.

5.5.3 MAINTAINING THE POSITION OF THE ASSESSEE: WHAT HINDERED AGENTIC POSITIONING?

As Studies III and IV noted, not all of the students taking part in summative self-assessment were agentially repositioning themselves. Furthermore, Studies I and II proposed that some student subgroups, drawing on a surface approach to learning, reported a low level of self-efficacy; these findings might indicate non-agentic positioning processes on a broader level. Through a discursive-deconstructive reading, these findings were reinterpreted. The

position of the assessee was not contested by all the students, and some students even further strengthened this positioning by objectifying themselves as the receivers of external assessment procedures (cf. Boud & Falchikov, 2006; Saari, 2011; Torrance, 2000). Thus, the “natural” position of the assessee was even further naturalised. Here, the main findings on how the position of the assessee was maintained are introduced.

As Study IV implied, some students saw the summative self-assessment model as another mechanism of governing and control (cf. Kasanen & Rätty, 2000). Arguably, the prevalent assessment culture cannot be substantially disrupted through only one university course implementation. This highlights Tan’s (2004) and Raaper’s (2019) notations of students as the negotiators of power in assessment. The position of the assessee was prevalent and supported by many structural discursive practices. The students’ perceptions painted a picture of mathematics assessment as test-taking, and the students were largely positioning themselves as non-agentic test-takers. This could be seen in Study III that raised concerns about the lack of agency of students in the summative self-assessment group.

Here, this “lacking agency” is reframed as the government of the self that summative self-assessment encouraged. Following Kasanen and Rätty (2002) and Tan (2004), it is implied that the summative self-assessment disciplined rather than empowered students by losing the *self* in self-assessment. This was evident in those students’ accounts in Study IV who thought that their conceptions of their own learning were not as sovereign as the feedback offered to them by the digital learning environment and by the student tutors. Summative self-assessment further constructed the position of “the assessee” by seemingly enabling freedom to students, but then guiding them through a self-assessment process in the desired way; if the assessee did not meet this proper pathway of deep learning and high performance (as promoted in Studies I and II), further governing towards the desired studying practices was needed. Study III identified self-monitoring of one’s own studying (which might fit the favourable goal of “self-regulation”) as a sign of lacking agency, which is a striking evidence of the government of the self. The government of the self was perhaps most evident in one student’s interview in which they described their refusal to self-grade their own work since the teacher would know their skills better. In a course with over 400 students this statement is arguably false!

In Study IV, the analysis of epistemological power revealed that the position of “the assessee” was constructed before the students took part in the summative self-assessment process - and quite probably long before their studies in higher education. In Studies II, III and IV, a frequent theme in the qualitative datasets was that summative self-assessment was considered as new, weird and even radical. What was identified through the discursive-deconstructive reading of the findings of Studies III and IV was the discourse of naturalising mathematics assessment as something naturally legitimate rather than socially constructed. The students largely framed self-assessment

as an unnatural process in mathematics; in Study IV, many students suggested that self-assessment might suit other disciplines better than mathematics. The government of the self was identified as students hoped for external examinations to validate their knowledge. There were optional digital tests about the mathematical content offering automatic feedback for those students who wished to test their own skills during the course; yet, these resources were widely underutilised. Maybe this was because these optional tests were not deemed to be sovereign - a proper examination was needed for that. These findings underline that the processes of “empowerment” and “promoting agency” do not happen in a vacuum. As the students in the summative self-assessment group had surely been subjectified as non-agentic assesseees already before their attendance of the summative self-assessment model, it was not simple for everyone to take part in technologies of self and self-subjectify oneself in an agentic way.

5.5.4 STUDENT POSITIONING IN THE RESEARCH PROJECT

So far, the deconstruction process has addressed student positioning in relation to assessment, based on the findings of Studies I-IV. However, the deconstruction process would be incomplete without deconstructing student positioning in the research articles themselves. Furthermore, this final deconstruction process is needed to highlight the ruptures between the various theoretical stances and epistemological premises, especially between Studies I-II and Studies III-IV. In a way, this final deconstruction of student positioning contests the approach of theoretical triangulation that this doctoral thesis has undertaken, and the opportunities and hindrances such an approach offers for disruptive research.

The relevance for Studies I and II was understood by the researcher to map out the broader picture of student positioning; whether agentic negotiation of students’ positioning has happened on a broader scale during large mathematics courses with hundreds of participants. This was indicated through measurements of deep and surface approaches to learning, self-efficacy and course achievement; Study II also drew on students’ open answers. All these quantitative instruments are largely utilised and statistically validated in the field of higher education research (e.g., Parpala & Lindblom-Ylänne, 2012), which justified their usage and likely contributed to enabling the publication of Studies I and II. However, the epistemological premises of these positivist endeavors greatly differ from the discursive approach taken in the synthesis.

While the intent was to utilise the quantitative datasets as “simple” indicators of positioning processes, and therefore as approaches that should be deepened through socio-cultural approaches in Studies III and IV, the deconstructive approach highlights that these studies do not simply provide knowledge about positioning but produce it. In Studies I and II, students were

positioned as “the assessees” through the research designs that rendered them as their scores in the research instruments. For example, Webb (1997) deconstructed deep and surface approaches to learning by bringing forth the individualising origins of this research tradition. Webb argues that the power is wielded as the deep approach to learning is seen as “the truth”, tying the notion with the Western Enlightenment tradition. While Studies I and II both drew on profiling methods not to reproduce the binary opposition of deep and surface approaches, neither aimed to disrupt that very opposition and the value imbalances it constructs. Webb’s criticism has already been addressed many times before (most notably in Entwistle, 1997), yet it reminds us that interpreting the quantitative findings of this doctoral thesis as an indicator of student positioning processes is not without its problems. Unfortunately, the “problems” in this thesis are quite substantial as they touch the very epistemological premises of Studies I-IV.

Studies III and IV drew on socio-cultural theories to deepen the findings of Studies I and II. Both studies explicitly distanced themselves from the individualistic approaches of Studies I and II, as they moved beyond seeing agency as power as something that resides in individuals. It should be noted that the opposition of the researchers and participants - or the ones being researched - was not disrupted. Both studies used interview data for their analyses. The students provided the data that the researchers analysed, and the positions in this process were uncontested. When it comes to research methodologies, the position of the participants (with its heavy resemblance to the position of the assessee) was maintained across Studies I-IV, no matter how much Studies I-IV aimed to promote the students’ role as active negotiators of agency and power. While this doctoral thesis and its substudies concern self-assessment, students’ self is notoriously absent in all these final research artefacts.

Webb argued that what is left outside the opposition of deep and surface approaches is the Other (1997, p. 210); the unthinkable that goes missing as the instrument produces information only about the known, *the thinkable*. As Studies I-IV have produced knowledge about subject positioning in relation to summative self-assessment, there is a need to investigate the Other in this process. First, the uncontested opposition of active/passive was present in all of the studies. Reflecting Webb’s (1997) critique on the deep and surface approaches to learning, students were steered towards the desired state of “activity”. This was conducted through active learning environments (Study I), active engagement (Study II), active self-assessment practices (Study III) and active agency (Study IV), for instance. Passivity needed to be prevented (Study II), and passive students were steered towards activity through the design of summative self-assessment. The opposition of activity/passivity was not conceptualised nor elaborated, and the desirable role of activity was naturalised; even when the position of the assessee was disrupted, students were steered towards activity. Second, while Studies I-IV all contributed in understanding summative self-assessment in the context of undergraduate

mathematics education, none of them reached beyond *the individual*; students were studying for themselves, not for their communities. The assessee is most certainly an individual, yet disrupting this position did not disrupt the individualistic premises of self-assessment.

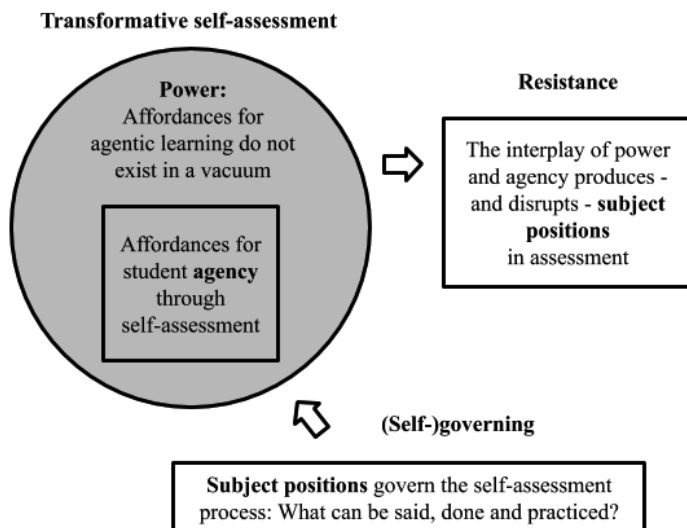
To sum up, the discursive-deconstructive reading of Studies I-IV has aimed to synthesise the findings of this doctoral thesis. Here, synthesis has referred to both building up smooth connections between the studies and highlighting the important ruptures between them. These ruptures are an important finding of this doctoral thesis. Finally, this synthesis is utilised to formulate the main contribution of this thesis: the concept of transformative self-assessment.

6 BEYOND DEEP SHIFT: TRANSFORMATIVE SELF-ASSESSMENT

Based on the discursive-deconstructive reading conducted for Studies I-IV, a novel theoretical concept is constituted: the concept of *transformative self-assessment*. This includes a future-driven element (Tan, 2007, 2008), while also acknowledging Allen's notion of resistance (2011). Hence, transformative self-assessment does not only foster self-reflection, but aims to disrupt the power relations of assessment through opening up a reflective space for students' agentic negotiation of their positioning. Transformative self-assessment is defined as consisting of practices that foster critical self-reflection yet also include the element of resistance by offering practical tools for self-transformation (Figure 3). Therefore, transformative self-assessment acknowledges both the viewpoints of agency and power.

While the formative self-assessment model was able to promote high quality studying (see the findings of Studies I and II), only the summative self-assessment model was connected with students' reflective self-transformation. Earlier higher educational research has largely tried to employ self-assessment to promote a deep approach to learning (Baeten, Dochy, & Struyven, 2008; Gijbels & Dochy, 2006; Rust, O'Donovan, & Price, 2005; Struyven et al., 2006) and student agency (Bourke, 2018; Milne, 2009; Taras, 2016). Transformative self-assessment does not aim to foster a deep approach to learning or to empower students, but to construct reflective spaces for students to renegotiate their positioning; in teacher-driven assessment cultures, this means disrupting the dominating position of "the assessee". It is argued that earlier studies on self-assessment in higher education have largely neglected the second part of Allen's (2011) notion of autonomy as resistance against disciplinary power and governmentality: the idea of self-transformation. Maybe this is why it has been shown that inducing "deep shifts" might be unsuccessful (Struyven et al., 2006) or even impossible (Haggis, 2003); if students' positioning as assessees is not disrupted, self-assessment is not transformative, and "the assessee" still remains naturalised. This doctoral thesis supplements previous contributions in the field, as the Foucauldian framework for positioning offers a novel perspective to understand self-assessment as a socio-cultural practice.

Figure 3 A summary of the concept of transformative self-assessment and its interconnections with the notions of agency, power and subject positioning.



It should be emphasised that transformative self-assessment is not a neutral concept, but a political and a disruptive one. Researchers willing to utilise it will need to take part in technologies of self and position themselves in the political field of assessment and grading policies (cf. Bagger, Björklund Boistrup, & Norén, 2018; Brunila & Valero, 2018). Therefore, transformative self-assessment does not only aim to promote agentic renegotiation of students' positioning but researchers' as well. Through technologies of the self, assessment researchers can become aware of the positioning processes in their field. Moreover, the practical side of the concept is crucial; what is considered to be disruptive practices varies according to the context. Researchers wishing to utilise the notion of transformative self-assessment should be in the frontline in designing disruptive self-assessment practices. In examination-driven contexts such as undergraduate mathematics education, it is not only crucial but *ethical* to understand and promote agentic student positioning. This does not happen through simple "empowerment practices" but through careful research design, scientific rigour and by strengthening the connection between research and practice.

In contexts where transformative assessment practices are not common, scaffolding systems are needed to support students in their agentic positioning. How this is best conducted, and with what kind of resources, is highly discipline-specific. In the summative self-assessment model as depicted in this doctoral thesis, students were offered support in various ways: self-assessment was practiced formatively and constructive feedback on the

process was offered, for instance. Also, the students could access an open learning space with available support anytime that suited their schedule (see Rämö et al., 2019). It can be hypothesised that in mathematics education a relatively large amount of support for agentic studying is needed, as the context largely positions students as assessees (Atjonen et al., 2019; Iannone & Simpson, 2011). As this doctoral thesis has shown, resistance is possible with larger class sizes as well, as long as the self-disciplinary grading processes are challenged. Identifying the various affordances of technology for the purposes of transformative self-assessment offers an interesting trajectory for future research.

The summative self-assessment model was an attempt to implement transformative self-assessment in practice. Obviously, the experiment was not a complete success. Rather, it reflected what one of Peter Filene's students (1969) noted about Filene's experiment on self-grading that was, in a lot of ways, similar to the one reported in this doctoral thesis: "You have a good system but lack a perfect society in which to use it." (p. 454) Indeed, the concept of transformative self-assessment, with its roots in the socio-cultural assessment literature, aims to shift the focus of research and practice from students' psychological processes into the structures of assessment. Already twenty years ago, Torrance (2000) argued that educators cannot close their eyes to the social control and self-discipline mechanisms of assessment. Similar issues have been raised for decades in relation to self-assessment (Kasanen & Rätty, 2000; Tan, 2004, Milne, 2009). Back in 1969, Filene stated that "one instructor cannot blithely try to make his courses a version of pedagogical utopia and at the same time use the symbols employed and defined in other ways by the non-utopian outer world." (p. 455). Indeed! The concept of transformative self-assessment calls for political actions to actively change the structures of that "non-utopian outer world". Tannock (2017) argues that assessment researchers should frame their work as political since unjust grading and assessment methods characterise higher education. Transformative self-assessment, based on the notion of resistance, builds on this call as it promotes assessment practices that challenge the power relations of grading and assessment by reaching beyond students' positioning as "the assessees".

7 SELF-REFLECTION OF THE RESEARCHER

A doctoral thesis about self-assessment would be incomplete without the author's own self-reflection about the process (unfortunately, doctoral theses cannot be self-graded!). While implementing a section for limitations is a common practice in educational research, the discursive-deconstructive reading requires deeper, systematic technologies of the self concerning the positioning of the researcher (cf. Foucault, 1998). This process is supplemented with reflection of the transferability of the findings and the limitations of the study. In this section, I intentionally change to first-person language to highlight the subjectivity in my own positioning.

7.1 MY POSITION AS A RESEARCHER, AND HOW IT HAS BEEN DISRUPTED

First of all, throughout conducting this doctoral study, my own positioning has been “a researcher” rather than a teacher, as all the data was collected from courses taught by someone else. This is crucial to note since this doctoral thesis is based on a practical assessment experiment. Yet, I would position this thesis in the intersection between theoretical and practical, as transformative self-assessment covers both aspects. Furthermore, this doctoral thesis represents the educational settings of higher education, mathematics education and assessment, and self-assessment in particular, just to name the most obvious ones; the approach taken has been multidisciplinary. As technologies of power have been identified to restrict the interaction between different educational disciplines (Bagger, Björklund Boistrup, & Norén, 2018; Brunila & Valero, 2018), the process of conducting this doctoral thesis has positioned me as a scientific nomad whose role is both to build bridges and emphasise ruptures where smooth connections cannot be built. Through discursive-deconstructive reading I have aimed to connect the qualitative and quantitative, theory and empiricism, research and practice - and reach beyond these dichotomies.

My own positioning as a specific kind of researcher has faced multiple disruptions throughout the process. I started my scientific career through a positivist lens, utilising quantitative measurements (Studies I and II). Therefore, I have actively taken part in those categorisation and normalisation processes that Foucault and those carrying on his work have identified and criticised (myself among them). As a researcher, I have an ethical responsibility to monitor how these findings are interpreted and applied. The first rupture in the research process emerged as I needed to explain the quantitative findings through the student interviews. While deepening

statistical findings through qualitative data is a common practice in educational studies, I soon realised that *deepening* my understanding would require me to reach beyond the positivistic premises of Studies I and II. This epistemological rupture resulted in the investigation of agency and power in Studies III and IV. This disruption of my doctoral candidacy has had concrete consequences; already as a young researcher I have come to notice the lack of discussion between the positivist and sociocultural approaches to understanding the complexity of self-assessment. As a referee from a journal that rejected Study IV put it: “The manuscript is not technically solid at all: no modeling, no theoretical analysis, and no quantitative results, no figures, no tables. It should be rejected directly.” Teaching in Higher Education, the critical journal where the article was published some months later, accepted the very same manuscript with rather minor revisions. This example is obviously anecdotal, but might reflect the silos of researchers in higher education. Working with this doctoral thesis at a mathematics department has further positioned me as a scientific nomad who does not exactly fit in.

Another disruption occurred while I was synthesising the findings of Studies I-IV through the discursive-deconstructive reading. Finally, I have come to position myself as a political researcher, as compared to *apolitical*. As Brunila and Valero (2018) note, the position of critical scholars is often challenged and marginalised (see also Bagger, Björklund Boistrup, & Norén, 2018); at the very beginning of my career, I have already faced empirical evidence of these phenomena. Reimagining the position of the assessee as constructed in both mathematics and higher education is more important than ever in the era of “21st century skills”. What could be more political than further constructing this non-agentic position? This is the time for assessment researchers to promote change (cf. Tannock, 2017) and disrupt those practices that hinder the development of students’ agency to unleash everyone’s mathematical potential for communities to use. Following Tannock (ibid.), I invite scholars in both mathematics and higher education to join the frontline of this mission. Brunila and Valero (2018) write about the cynicism of critical scholars; personally, I feel empowered and, above all, hopeful.

7.2 TRANSFERABILITY OF THE FINDINGS

Neither the quantitative (Studies I and II) or the qualitative (Studies III and IV) approaches utilised in this doctoral thesis seek the generalisability of the findings. However, the transferability of the findings can be discussed. Even though the discursive-deconstructive reading admits its specificity on its socio-cultural context, the approach is aimed towards a concrete impact in the field of education - both in terms of theory and practice. First, the findings of this doctoral thesis are transferable to the field of mathematics education. Assessment of mathematics is globally strongly built on teacher-driven

practices (e.g., Nortvedt & Buchholtz, 2018) just as it is in Finland (Atjonen et al., 2019). In teacher-driven assessment cultures like the one of mathematics, resistance through disruptive practices such as self-assessment might produce similar kinds of effects on students' positioning. Similar subject positioning processes might be identified from mathematical contexts beyond undergraduate education where classroom assessment culture is examination-driven. Contrasting the subject positions that mathematics assessment produces for students in various educational contexts offers an interesting subject for future research.

One needs to be careful while considering the transferability of these findings in the field of higher education. The Finnish context is based on low-stakes assessment; academic freedom is provided for teachers in the legislation (Finnish Universities Act, 558/2009). This is reflected in practice: Finland ranks highly in international comparisons for university teachers' autonomy on their own teaching methods (Nokkala & Bladh, 2014). At the same time, undergraduate mathematics provides an examination-driven context for this doctoral thesis. Based on this, I imply that the findings might be transferable to other teacher-driven assessment cultures in higher education. It might be that in contexts where assessment is generally based on student-centred practices (such as peer- and self-assessment), less support is needed for agentic studying; the students might already have been prompted to reflect on their positioning through disruptive assessment practices. Finally, this doctoral thesis has developed theoretical understanding of the interplay between agency and power by introducing the concept of transformative self-assessment. This concept should be applied in various higher educational contexts with both high- and low- stakes assessment.

7.3 LIMITATIONS AND ETHICAL CONSIDERATIONS

The research process of this doctoral thesis followed the ethical principles of the mathematics education research group (Hannula, Lahdenperä, & Nieminen, 2018). These principles reflected the ethical instructions of the Finnish National Board on Research Integrity, and they were constructed with the assistance of the Centre for University Teaching and Learning at the University of Helsinki. Writing, revising and implementing these ethical principles was a substantial part of my overall learning process as a doctoral candidate. Voluntary participation, informed consent, anonymity, confidentiality and secure data storage formed the basis for the ethical considerations conducted through the research project.

Studies I-IV were all based on a rather simple design in which all the data was collected after the course. This lack of longitudinal follow-up data is an obvious limitation that should be considered while interpreting the findings. The data allowed the researcher to deconstruct the subject positioning

processes only through interview and survey data; these might be distorted from students' lived experiences. Furthermore, the quantitative measurements used no pre- and post-data. The research design in Studies I, III and IV might raise ethical concerns, as the two self-assessment groups were graded through different methods. Another limitation for both the research designs (Studies I, III-IV and Study II) is the lack of participatory approaches. The participants had no chance to comment on the interpretations made by the researcher. The students are only seen through that position: they are subjectified as the participants in a study about agency and power. This is just as unflattering as it sounds.

Researcher triangulation was utilised in studies I, II and III in all parts of the research process (Denzin, 1978; Van Drie & Dekker, 2013). The researchers discussed the findings during the analysis process, and the qualitative data analysis validation was reported (Studies III and IV). In Study III, the "validity check" for the analytical path was an important part of the process of understanding the interplay of agency and self-assessment, as it allowed the researchers to further build their theoretical understanding in conversation with the data. Study IV requires a closer examination since it was single-authored and I conducted the analysis myself. However, the elaborative coding method (Auerbach & Silverstein, 2003; Saldaña, 2016) utilised in the study was not exactly aiming for objective analysis, in positivistic means. On the contrary, the analytical process built on the researcher's knowledge about the theoretical frameworks of power. In this study, transparent reporting of the findings and how they were interpreted formed the base for "validation" of the data analysis that drew on elaborative coding.

Furthermore, the discursive-deconstructive reading (Ikävalko & Brunila, 2019) as the main approach of this doctoral thesis is only conducted through my own positioning (see Section 7.1). In this thesis, I have been transparent in terms of methodological and theoretical choices made - while criticising earlier studies on self-assessment for not always opening up their analytical paths (see studies Studies III and IV). Finally, peer debriefing (Given, 2008) was widely utilised during the whole process of conducting this doctoral thesis and its substudies. The preliminary findings of Studies I-IV have been presented during research visits at Deakin University, The University of Tasmania, San Diego State University and The University of Melbourne. Personally, I think these research visits have been the most fruitful part of my doctoral candidacy, and the final deconstruction process would not have looked the same without them.

The quantitative methodology of studies I and II demands a closer investigation. Self-reported research questionnaires about students' own experiences have been deemed unreliable as external factors might affect students' responses (Cohen, Manion, & Morrison, 2007). Nonetheless, the instruments were validated in the context of Finnish higher education (Herrmann, Bager-Elsborg, & Parpala, 2017). Achievement data was used to indicate learning in both research designs of this thesis. Calculating the

average score of the course assignments is a simple way of conceptualising achievement. No other course data, such as engagement with the digital learning environment, was utilised.

The qualitative data collection was mainly conducted by me, while another researcher from the Digital Self-Assessment project interviewed three of the students. The interviews were transcribed by a professional. Through a discursive-deconstructive reading it is possible to recognise the power relations that are always present in interview situations. For instance, it might be that some maladaptive agentic orientations (Study III) were not identified since students were aware of their positioning both as “the student” and as “the participant”, both of which might have restricted what could be said and how.

8 IMPLICATIONS

8.1 IMPLICATION FOR THEORY

First, as its main contribution, this doctoral thesis has constituted the concept of transformative self-assessment. The research communities on self-assessment in higher education, and on assessment in undergraduate mathematics, are invited to engage with this concept.

Overall, this doctoral thesis has positioned itself as a socio-cultural study on self-assessment, calling for future research on self-assessment to build further understanding of the interplay of self-assessment and its socio-cultural contexts. This might be especially crucial in examination-driven contexts such as the one of undergraduate mathematics education as described in this thesis (Iannone & Simpson, 2011). Socio-cultural frameworks could supplement the vast amount of psychological and cognitive literature to understand self-assessment in relation to its educational context (cf. Andrade, 2019). While the psychological perspectives are certainly crucial to conceptualise self-assessment, the lack of socio-cultural approaches might hinder the possible interpretations of findings. Take, for instance, the study on self-grading by Tejeiro and colleagues (2011) that has been framed as an example to warn against using self-grading practices in educational settings (e.g., Andrade, 2019). In Tejeiro et al's study it was found that the students overestimated their own grades when they had a chance to self-grade 5% of their own course mark. The students were held responsible. While the grade inflation in Tejeiro and colleagues' study certainly tells something about the students themselves, the role of the educational context and the social factors of learning would have deepened the implications of this study. For example, it could have been asked why the learning environment encouraged maladaptive rather than adaptive form of agency. Furthermore, concerns arise if the domination of individual perspectives renders structural issues of assessment into personal and psychological (cf. Ikävalko & Brunila, 2019). For instance, framing the issue of assessment that hinders agency as students' lack of cognitive abilities hardly fosters the design of assessment practices in contemporary higher education.

If socio-cultural approaches are employed, scientific rigour should not be forgotten. Concepts such as agency and power need to be tied into strong theoretical frameworks; as noted in Study III, this is not always the case in terms of self-assessment in higher education. Furthermore, these theoretical frameworks should be aligned with carefully designed methodologies. For example, future qualitative research on agency and power should develop theory-driven analysis methods rather than leaning on data-driven analyses (see Study III). As argued in Study IV, the theoretical frameworks for addressing structural issues of agency and power in assessment, and self-

assessment in particular, already exist; future studies are encouraged to engage with these educational theories in multidisciplinary ways.

Furthermore, it is suggested that the dialogue between the psychological and socio-cultural fields of assessment research needs to be fostered. In this doctoral study, the Foucauldian concept of subject positioning was used to synthesise studies drawing on both individual and socio-cultural epistemological premises. Discursive-deconstructive reading (Ikävalko & Brunila, 2019) offered conceptual tools for this. Without multidisciplinary approaches, self-assessment research might result in further silos without enough discussion between fields with different methodological and even epistemological premises. This process might limit researchers' agency in interpreting their findings through multidisciplinary lenses (cf. Brunila & Valero, 2018).

Leach and colleagues (2001) questioned the “neutrality” of the mere act of empowerment. Here, similar critical perspectives are called towards research designs on self-assessment. The notions of agency and power are strong theoretical frameworks for understanding researchers' actions as much as they are for students'. Based on the findings of Studies I-IV and their synthesis, it is implied that the lack of holistic and critical approaches towards assessment research designs might have in part prevented alternative assessment methods from resulting in deep shifts (cf. Haggis, 2003). Just as Haggis suggested, a deep approach to learning might not even be possible to induce for students. In this doctoral thesis it is argued that this could result from insufficient understanding of student positioning; if students are positioned as “the assessee” who should foster their learning and studying only through that position, the assessment practice in hand is hardly transformative. As shown in this thesis, a discursive-deconstructive reading enables a comprehensive approach to understanding one's own positioning as a researcher through technologies of self (Bagger, Björklund Boistrup, & Norén, 2018). This offers a powerful tool for future research on self-assessment to understand the positioning processes of the field. What kinds of research topics are fostered and funded, and what kinds of voices are silenced?

8.2 IMPLICATIONS FOR METHODOLOGY

Following Panadero and colleagues (2016; see also Andrade, 2019), this doctoral thesis calls for longitudinal research designs to understand subject positioning in self-assessment. The longer-term effect of transformative self-assessment practices on students' positioning would offer an interesting topic for future research. Would reflective spaces open up frequently for agentic negotiation of student positioning if transformative self-assessment practices would be used longitudinally? Could transformative self-assessment even become the *normal* way to assess? Longitudinal studies focusing on both

methodological and theoretical triangulations would offer deeper insight on the interplay of self-assessment and subject positioning.

As noted before, this doctoral thesis lacks both participatory approaches and student voice (cf. Taylor & Robinson, 2009); both of these are recommended for future socio-cultural research on self-assessment. The use of students as partners in the assessment research processes has been rare. Neither of the recent literature reviews on self-assessment research identify participatory approaches as a possible future trajectory (Andrade, 2019; Panadero et al., 2016). Instead, Panadero and colleagues (ibid.) note that it is important for future research to understand that not all students benefit from the same self-assessment practices in a similar way. Profiling and clustering methods offer a novel approach for future research to understand student subgroups in self-assessment; yet, even more crucial would be to deepen the understanding of how various student subgroups could participate in the design and research processes concerning self-assessment. This is especially crucial in examination-driven contexts where students have preferred traditional assessment practices, implying their positioning as “assesseees”.

Finally, following the theoretical implications above, this doctoral thesis encourages self-assessment researchers to position themselves in the field of assessment and politics - and to reflect on their research methodologies. As Tannock (2017) notes, assessment research has built a vast amount of knowledge on what kinds of pedagogical practices support learning, yet assessment researchers rarely take part in broader political discussions about assessment and grading in higher education. Novel openings through disruptive research designs might be needed if the field of self-assessment is to be more than “a nice idea and second runner in assessment” (Panadero et al., 2016, p. 824). Transformative self-assessment is by definition based on concrete actions that enable resistance; examining transformative research designs on self-assessment offers a challenge for future research. Novel approaches would be needed in terms of method triangulation as well. This doctoral thesis heavily drew on student interview data; future research could search for novel data construction methods. Digital technologies and student participation might offer ways to conduct novel forms of self-assessment for data construction as well.

8.3 IMPLICATIONS FOR PRACTICE

The teachers in examination-driven assessment cultures, such as in undergraduate mathematics education, should critically examine and contest the positions their students are enabled to undertake through their assessment practices. The findings of Studies III and IV underlined students’ non-agentive positioning as assesseees that they had constructed during their earlier mathematics studies. Based on these findings, educators in higher education,

and in undergraduate mathematics education in particular, are asked to take part in technologies of self to critically examine their assessment practices - and the underlying assumptions of these practices. As Tan argues (2004), teachers' self-reflection is crucial, especially given the complexity of promoting students' agency within the power relations of the specific socio-cultural-political context. One way of addressing the issues of assessment, power and agency in contexts where teacher-driven assessment dominates is structural support and training for teachers for developing their assessment methods through active self-reflection.

The notion of transformative self-assessment offers an interesting perspective for teachers in higher education. Educators wishing to promote student agency need to implement concrete support mechanisms in their assessment environments that students can utilise for self-transformation. The question of what those support mechanisms are is strictly context-specific. The findings of this thesis have shown that resistance through self-assessment is possible in the context of large courses as well, as long as the support system is carefully designed. If assessment practices are not designed for self-transformation, it might be that assessment becomes a factor that hinders agentic studying; it is argued that this issue is not only unethical but *political* as well. In this case assessment might hamper the educational benefits of instructional practices aiming to promote agentic or higher quality studying. On a practical level, this might even lead to waste of resources.

Finally, this doctoral thesis holds broader pedagogical implications for higher educational institutes and faculties. For example, the Finnish Universities Act asks the universities to educate students to "serve their country and humanity at large" (558/2009) rather than only teaching them an accustomed set of skills. Similar ideas can be found in the Strategic Plan of the University of Helsinki (2017-2020): "*Universities must be able to develop education further in accordance with the competence demands and labour market needs of the future. Today's students will be solving tomorrow's challenges. – – Both electronic learning environments and customisable teaching facilities will support learners as active agents.*" How could assessment practices support these kinds of goals? Based on the finding of this doctoral thesis, novel assessment and grading policies are called for that enable students to agentially reposition themselves (cf. Boud & Falchikov, 2006). Here it has been shown that summative self-assessment enabled agentic student positioning, but only because it was used in a transformative way.

9 FINAL CONCLUSIONS

This doctoral thesis has introduced the concept of transformative self-assessment to conceptualise self-assessment as a disruptive practice. In this doctoral thesis, the “natural” positioning for students as the assessees, or as the non-agentic receiver of assessment, was disrupted. What differentiates transformative self-assessment from other self-assessment practices is its connection with *resistance*. In Edgar Allan Poe’s “The System of Doctor Tarr and Professor Fether” (1845), the starting point and the final closure for this thesis, resistance took the form of tar and feathers. In the story, resistance was only made possible by enabling agency for the lunatics, who then took over - and completely shifted what is meant by lunacy in the first place. Poe’s story offers a playful way of conceptualising transformative self-assessment. It might be that transformative approaches, such as self-grading, would lead to “the notorious ‘I give myself an A’” (Andrade & Du, 2007, p. 160), and not to that deep shift (Haggis, 2003) we were hoping for. That is the thrill of empowerment, which has been advocated as the goal of self-assessment in higher education (Milne, 2009; Taras, 2016). However, only enabling the non-agentic position of “the assessee” through assessment is not an apolitical act but political; further, the position is unnatural rather than natural. This is why transformative practices are needed in contexts such as undergraduate mathematics education.

Simple claims such as how self-grading should not be used in higher education are certainly not sufficient to theorise the complex interrelations between grading, assessment and learning, let alone to guide the development of assessment that would promote agency. So, finally, the concluding remark of this doctoral thesis is not to “promote summative self-assessment as a viable assessment method”, as a warm-hearted colleague once put it. Rather, to understand the complexity of self-assessment, I call for novel, ambitious and disruptive research to build bridges between practice and theory, quantitative and qualitative, positivist and discursive; and to highlight ruptures where bridges cannot be built. In Poe’s story the concept of *lunacy* was examined from the perspective of societal norms rather than the perspective of individual differences, turning the very premises of what it means to be *mad* upside down; similarly, the multifaceted nature of self-assessment needs to be understood through a variety of approaches that includes the investigation of socio-cultural structures. Such investigations should shake up our understanding of certain kinds of mathematics learners, and highlight how assessment *creates* weak students, overachievers, the ones who cannot self-assess, the surface learners, and so on. In the end, that is exactly what assessment should aim for: to highlight the complexity of learning and studying rather than aiming to simplify these multifaceted phenomena.

REFERENCES

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* 52: 317–32.
- Alhanen, K. (2018). *Practices and thought in Michel Foucault's philosophy*. Helsinki: Books on Demand.
- Allen, A. (2011). Foucault and the politics of our selves. *History of the Human Sciences*, 24(4), 43–59.
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4: 87. doi:10.3389/educ.2019.00087
- Andrade, H. L., & Brown, G. T. L. (2016). Student self-assessment in the classroom. In G. T. L. Brown & L. R. Harris (Ed.), *Handbook of human and social conditions in assessment* (pp. 319–334). New York: Routledge.
- Andrade, H. L., (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Ed.), *Handbook of formative assessment* (pp. 102–117). New York: Routledge.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48(1), 12–19.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32(2), 159–181.
- Arribas-Ayllon, M. & Walkerdine, V. (2008). Foucauldian discourse analysis. In C. Willig & W. Stainton-Rogers (Ed.), *Handbook of qualitative research in psychology* (pp. 91–108). London: Sage.
- Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students' deep and surface approaches to learning in higher education. *Educational Psychology Review*, 29(2), 205–234.
- Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kampi, P., Rumpu, N., Hietala, R., & Immonen, J. (2019). *"Että tietää missä on menossa". Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa*. Helsinki: The publications of the Finnish Education Evaluation Centre.
- Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York: New York University Press.
- Bacchi, C., & Bonham, J. (2014). Reclaiming discursive practices as an analytic focus: Political implications. *Foucault studies*, 17, 179–192.
- Baeten, M., Dochy, F., & Struyven, K. (2008). Students' approaches to learning and assessment preferences in a portfolio-based learning environment. *Instructional Science*, 36(5–6), 359–374.
- Bagger, A., Björklund Boistrup, L., & Norén, E. (2018). The governing of three researchers' technologies of the self. *The Montana Mathematics Enthusiast*, 15(1–2), 278–302.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175–1184.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, 9(3), 75–78.

- Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education*, 36(6), 671-687.
- Biesta, G., Priestley, M., & Robinson, S. (2015). The role of beliefs in teacher agency. *Teachers and Teaching*, 21(6), 624-640.
- Biesta, G., & Tedder, M. (2007). Lifelong learning and the ecology of agency: Towards a lifecourse perspective. *Studies in the Education of Adults*, 39(2), 132-149.
- Biggs, J. B. (1991). Approaches to learning in secondary and tertiary students in Hong Kong: Some comparative studies. *Educational Research Journal*, 6(1), 27-39.
- Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal*, 29(5), 623-637.
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Ed.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 14-26). London: Routledge.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529-549.
- Bourke, R. (2018). Self-assessment to incite learning in higher education: developing ontological awareness. *Assessment & Evaluation in Higher Education*, 43(5), 827-839.
- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444-457.
- Brown, G. T., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 2(1), 22-30.
- Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (ed.), *SAGE handbook of research on classroom assessment* (pp. 367-393). Los Angeles, CA: Sage.
- Brunila, K., & Valero, P. (2018). Anxiety and the making of research(ing) subjects in neoliberal academia. *Subjectivity*, 11(1), 74-89.
- Brunila, K. & Ikävalko, E. (2012). Vallan tunnistaminen siihen lamaantumatta: näkökulmia diskursiivis-dekonstruktiiiviseen lukutapaan. *Politiikka*, 53(4), 285-300.
- Burr, V. (2015). *Social constructionism*. London: Routledge.
- Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education*, 44(5), 705-714.
- Charteris, J., & Smardon, D. (2018). A typology of agency in new generation learning environments: Emerging relational, ecological and new material considerations. *Pedagogy, Culture & Society*, 26(1), 51-68.
- Charteris, J., & Thomas, E. (2017). Uncovering 'unwelcome truths' through student voice: Teacher inquiry into agency and student assessment literacy. *Teaching Education*, 28(2), 162-177.
- Clegg, S. R. (1989). *Frameworks of power*. London: Sage.

- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. London: Routledge.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons.
- Denzin, N. (1978). *Sociological methods: A sourcebook*. New York: McGraw Hill.
- DeSantis, L., & Ugarriza, D. N. (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research*, 22(3), 351-372.
- Diseth, Å. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality*, 17(2), 143-155.
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, 103(4), 962-1023.
- Entwistle, N. (1997). Reconstituting approaches to learning: A response to Webb. *Higher Education*, 33(2), 213-218.
- Entwistle, N., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.
- Evans, A. M. (2011). Governing student assessment: Administrative rules, silenced academics and performing students. *Assessment & Evaluation in Higher Education*, 36(2), 213-223.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Filene, P. G. (1969). Self-grading: An experiment in learning. *The Journal of Higher Education*, 40(6), 451-458.
- Foucault, M. (1990). *The history of sexuality: An introduction (Vol. 1)*. New York: Vintage Books.
- Foucault, M. (1988). Technologies of the self. In L. H. Martin, H. Gutman, & P. H. Hutton (Eds.), *Technologies of the self: A seminar with Michel Foucault* (pp. 16-49). London: Tavistock.
- Foucault, M. (1982). The subject and power. *Critical Inquiry* 8(4), 777-795.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison*. New York: Vintage Books.
- Foucault, M. (1991). Questions of Method. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality with two lectures and an interview with Michel Foucault* (pp. 73-86). Chicago, IL: University of Chicago Press.
- Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference? *Educational Studies*, 32(4), 399-409.
- Given, L. M. (Ed.). (2008). *The SAGE encyclopedia of qualitative research methods*. London: Sage publications.
- Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, 101(2), 336-342.
- Hanafin, J., Shevlin, M., Kenny, M., & Neela, E. (2007). Including young people with disabilities: Assessment challenges in higher education. *Higher Education*, 54(3), 435-448.
- Hannula, J., Lahdenperä, J., & Nieminen, J. H. (2018). *Matematiikan tutkimusryhmän tutkimuseettiset periaatteet* [The ethical principles of the mathematics education research group]. Acta generalia instituti mathematico-rationarii. <http://hdl.handle.net/10138/301728>

- Harris, L. R., Brown, G. T., & Dargusch, J. (2018). Not playing the game: Student assessment resistance as a form of agency. *The Australian Educational Researcher*, 45(1), 125-140.
- Haggis, T. (2003). Constructing images of ourselves? A critical investigation into 'approaches to learning' research in higher education. *British Educational Research Journal*, 29(1), 89-104.
- Hanson, F. A. (2002). How tests create what they are intended to measure. In A. Filer (ed.), *Educational assessment and testing: Social practice and social product* (pp. 67-81). London: Routledge Falmer.
- Heller, K. J. (1996). Power, subjectification and resistance in Foucault. *SubStance*, 25(1), 78-110.
- Herrmann, K. J., Bager-Elsborg, A., & Parpala, A. (2017). Measuring perceptions of the learning environment and approaches to learning: validation of the learn questionnaire. *Scandinavian Journal of Educational Research*, 61(5), 526-539.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Iannone, P., & Simpson, A. (2011). The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 30(4), 186-196.
- Iannone, P., & Simpson, A. (2015a). Students' preferences in undergraduate mathematics assessment. *Studies in Higher Education*, 40(6), 1046-1067.
- Iannone, P., & Simpson, A. (2015b). Students' views of oral performance assessment in mathematics: straddling the 'assessment of and 'assessment for' learning divide. *Assessment & Evaluation in Higher Education*, 40(7), 971-987.
- Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59(2), 243-258.
- Ikävalko, E., & Brunila, K. (2019). Coming to discursive-deconstructive reading of gender equality. *International Journal of Research & Method in Education*, 42(1), 33-45.
- Jay, J., & Owen, A. (2016). Providing opportunities for student self-assessment: The impact on the acquisition of psychomotor skills in occupational therapy students. *Assessment & Evaluation in Higher Education*, 41(8), 1176-1192.
- Kasanen, K., & Rätty, H. (2002). "You be sure now to be honest in your assessment": Teaching and learning self-assessment. *Social Psychology of Education*, 5(4), 313-328.
- Kennewell, S. (2001). Using affordances and constraints to evaluate the use of information and communications technology in teaching and learning. *Journal of Information Technology for Teacher Education*, 10(1-2), 101-116.
- Knight, P. T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education*, 27(3), 275-286.
- Lahdenperä, J., Postareff, L., & Rämö, J. (2019). Supporting quality of learning in university mathematics: A comparison of two instructional designs. *International Journal of Research in Undergraduate Mathematics Education*, 5(1), 75-96.

- Leach, L., Neutze, G., & Zepke, N. (2001). Assessment and empowerment: Some critical questions. *Assessment & Evaluation in Higher Education*, 26(4), 293-305.
- Lilja, M., & Vinthagen, S. (2014). Sovereign power, disciplinary power and biopower: Resisting what power with what resistance? *Journal of Political Power*, 7(1), 107-126.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778.
- López-Pastor, V. M., Fernández-Balboa, J. M., Santos Pastor, M. L., & Fraile Aranda, A. (2012). Students' self-grading, professor's grading and negotiated final grading at three university programmes: Analysis of reliability and grade difference ranges and tendencies. *Assessment & Evaluation in Higher Education*, 37(4), 453-464.
- Maciejewski, W., & Merchant, S. (2016). Mathematical tasks, study approaches, and course grades in undergraduate mathematics: A year-by-year analysis. *International Journal of Mathematical Education in Science and Technology*, 47(3), 373-387.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, 46(1), 4-11.
- Meijer, P. C., Verloop, N., & Beijaard, D. (2002). Multi-method triangulation in a qualitative study on teachers' practical knowledge: An attempt to increase internal validity. *Quality and Quantity*, 36(2), 145-167.
- Miles, M.B., Huberman, A.M., Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.
- Milne, C. (2009). Assessing self-evaluation in a science methods course: Power, agency, authority and learning. *Teaching and Teacher Education*, 25(5), 758-766.
- Morgan, C., Tsatsaroni, A., & Lerman, S. (2002). Mathematics teachers' positions and practices in discourses of assessment. *British Journal of Sociology of Education*, 23(3), 445-461.
- Mumm, K., Karm, M., & Remmik, M. (2016). Assessment for learning: Why assessment does not always support student teachers' learning. *Journal of Further and Higher Education*, 40(6), 780-803.
- Murphy, P. E. L. (2017). Student approaches to learning, conceptions of mathematics, and successful outcomes in learning mathematics. In L. Wood & Y. Breyer (Eds.), *Success in Higher Education* (pp. 75-93). Singapore: Springer Singapore.
- Naskali, P. (2003). Dekonstruktio uusien kysymysten avaajana kasvatustieteessä ja yliopistopedagogiikassa [Deconstruction opening up new questions in education and academic pedagogy]. *Kasvatus*, 34, 18-29.
- Nokkala, T., & Bladh, A. (2014). Institutional autonomy and academic freedom in the Nordic context—similarities and differences. *Higher Education Policy*, 27(1), 1-21.
- Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, 50(4), 555-570.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74-98.

- Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803-830.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813.
- Parpala, A., & Lindblom-Ylänne, S. (2012). Using a research instrument for developing quality at the university. *Quality in Higher Education*, 18(3), 313-328.
- Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., & Hirsto, L. (2010). Students' approaches to learning and their experiences of the teaching-learning environment in different disciplines. *British Journal of Educational Psychology*, 80(2), 269-282.
- Patton, C. (2012). 'Some kind of weird, evil experiment': Student perceptions of peer assessment. *Assessment & Evaluation in Higher Education*, 37(6), 719-731.
- Phillips, K. R. (2006). Rhetorical maneuvers: Subjectivity, power, and resistance. *Philosophy & Rhetoric*, 39(4), 310-332.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., and McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Post-Secondary Teaching.
- Prat-Sala, M., & Redford, P. (2010). The interplay between motivation, self-efficacy, and approaches to studying. *British Journal of Educational Psychology*, 80(2), 283-305.
- Raaper, R. (2019). Students as consumers? A counter perspective from student assessment as a disciplinary technology. *Teaching in Higher Education*, 24(1), 1-16.
- Rajala, A., & Kumpulainen, K. (2017). Researching teachers' agentic orientations to educational change in Finnish schools. In M. Goller & S. Paloniemi (Ed.), *Agency at work: An agentic perspective on professional learning and development* (pp 311-329). Cham: Springer.
- Reynolds, M., & Trehan, K. (2000). Assessment: a critical perspective. *Studies in Higher Education*, 25(3), 267-278.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353.
- Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231-240.
- Rämö, J., Reinholz, D., Häsä, J., & Lahdenperä, J. (2019). Extreme Apprenticeship: Instructional change as a gateway to systemic improvement. *Innovative Higher Education*, 44(5), 351-365.
- Saari, A. (2011). *Kasvatustieteen tiedontahto. Kriittisen historian näkökulmia suomalaiseen kasvatuksen tutkimukseen. Kasvatusalan tutkimuksia 55*. Jyväskylä: Suomen kasvatustieteellinen seura.
- Saldaña, J. (2016). *The coding manual for qualitative researchers*. London: Sage.
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

- Strong, B., Davis, M., & Hawks, V. (2004). Self-grading in large general education classes: A case study. *College Teaching*, 52(2), 52-57.
- Struyven, K., Dochy, F., Janssens, S., & Gielen, S. (2006). On the dynamics of students' approaches to learning: The effects of the teaching/learning environment. *Learning and Instruction*, 16(4), 279-294.
- Tan, K. H. K. (2008). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research & Development*, 27(1), 15-29.
- Tan, K. H. K. (2007). Conceptions of self-assessment: What is needed for long term learning? In D. Boud & N. Falchikov (Ed.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 114-127). London: Routledge.
- Tan, K. H. K. (2004). Does student self-assessment empower or discipline students? *Assessment & Evaluation in Higher Education*, 29(6), 651-662.
- Tannock, S. (2017). No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university. *Studies in Higher Education*, 42(8), 1345-1357.
- Taras, M. (2016). Situating power potentials and dynamics of learners and tutors within self-assessment models. *Journal of Further and Higher Education*, 40(6), 846-863.
- Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teaching in Higher Education*, 13(1), 81-92.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26(6), 605-614.
- Taylor, C., & Robinson, C. (2009). Student voice: Theorising power and participation. *Pedagogy, Culture & Society*, 17(2), 161-175.
- Tejeiro, R. A., Gomez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., & Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology*, 10(2), 789-812.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 14(3), 281-294.
- Torrance, H. (2000). Postmodernism and educational assessment. In A. Filer (Ed.), *Assessment. social practice and social product* (pp. 173-188). London, UK: Routledge/Falmer.
- Webb, G. (1997). Deconstructing deep and surface: Towards a critique of phenomenography. *Higher Education*, 33(2), 195-212.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Wilson, K., & Fowler, J. (2005). Assessing the impact of learning environments on students' approaches to learning: Comparing conventional and action learning designs. *Assessment & Evaluation in Higher Education*, 30(1), 87-101.
- Van Drie, J., & Dekker, R. (2013). Theoretical triangulation as an approach for revealing the complexity of a classroom discussion. *British Educational Research Journal*, 39(2), 338-360.
- Yan, Z., & Brown, G. T. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, 42(8), 1247-1262.

